No cover image available

Critical Appraisal of Epidemiological Studies and Clinical Trials (3rd edn)

Mark Elwood

https://doi.org/10.1093/acprof:oso/9780198529552.001.0001

Published online: 01 September 2009 **Published in print:** 22 February 2007 **Online ISBN:** 9780191723865 **Print ISBN:** 9780198529552

CHAPTER

1 The importance of causal relationships in medicine and health care

J. Mark Elwood

https://doi.org/10.1093/acprof:oso/9780198529552.003.01 Pages 5–18

Published: February 2007

Abstract

This chapter discusses the concept of causation. Topics covered include definition and types of causation, a direct test of causation (the randomized trial), and methods of counting events. Self-test questions are provided at the end of the chapter.

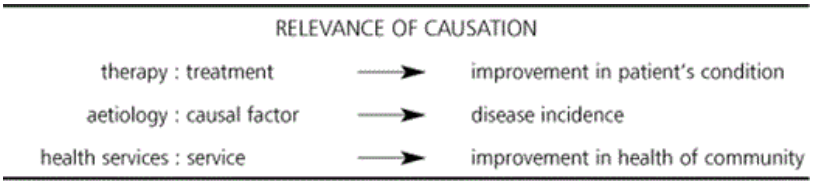
Keywords: causation, cause and effect, prevalence, incidence, critical appraisal

Subject: Public Health, Epidemiology

Collection: Oxford Scholarship Online

> *Intellectual progress is by no one trait so adequately characterized as by the development of the idea of causation.*
>
> —Herbert Spencer: The data of ethics, IV; 1879

The study of causal relationships is essential to medicine and health care (**Ex.** 1.1). In treating a patient, the decision to offer a treatment is based on the assumption that that treatment will *cause* an improvement in the patient's condition. The study of the causes and prevention of a disease involves determining which factors *cause* the disease to occur. In health service management, the decision to provide a certain type of service or facility assumes that it will *cause* an improvement in the health of the individuals or the community that it serves. On my desk today there are new studies on whether the prenatal environment leads to diabetes, assessments of new treatments for heart disease, malaria, and bladder cancer, and a discussion on whether clinical leadership improves the performance of health services. These articles essentially ask questions of causality. Does a specific treatment cause an improvement in the patient's condition? Do particular factors cause diseases to occur? Do different systems of health management cause improvements for the users of the service? Most critical issues in health care and most controversial issues depend on the assessment of whether a cause and effect relationship exists.



Ex. 1.1. The relevance of causal relationships in health care. Decisions on therapy, aetiology, or health services provision all make assumptions of cause and effect relationships

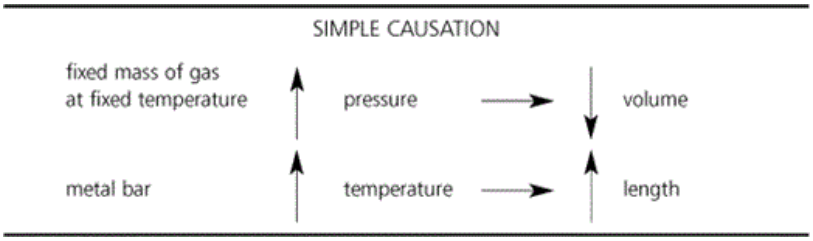
In this book, we will explore the methods available to us to test whether relationships are in fact causal, and therefore to decide whether the assumptions behind decisions relating to therapy, aetiology, and health service management are true or false.

This *critical appraisal* process is central to the related topics of *meta-analysis* and *evidence-based medicine*. The evaluation of what is done and what is not done in health care, in terms of its basis in scientific evidence, depends on correct judgements being made about causal relationships.

p. 6 We will concentrate on the critical appraisal of epidemiological studies and clinical trials. We will present a practical system for critical appraisal, which can be applied to any of the main types of study that assess relationships between an *intervention or exposure* and an *outcome*. The central issue is whether any association seen between the intervention or exposure and the outcome indicates a cause and effect relationship; if so appropriate action can follow. If a cause and effect relationship does not exist, the association must be due to other mechanisms; this conclusion has different practical consequences.

Definition of causation, and types of causation

To discuss causal relationships, we must have a definition of causation, and this definition will determine how causation can be demonstrated or disproved. The concept of causation often brings to mind only the extreme and limiting situation, which is the situation where a certain event *always and invariably* follows another event. This is a familiar notion because it is regularly observed in the physical sciences. For example, Boyle's law states that at a given temperature, the volume of a fixed mass of a gas is inversely proportional to its pressure. Thus a change in pressure results invariably and automatically in a corresponding change in volume (**Ex.** 1.2). The effect is instantaneous, can be replicated easily, and can be expressed as a simple mathematical relationship. Therefore there is little difficulty in accepting the notion that a change in pressure *causes* a change in volume. Similarly, there is little difficulty in the concept that applying heat to a metal bar *causes* it to expand; again there is an invariable and almost immediate relationship between the 'outcome', i.e. the change of length of the bar, and the 'exposure', i.e. the heat applied to the bar.

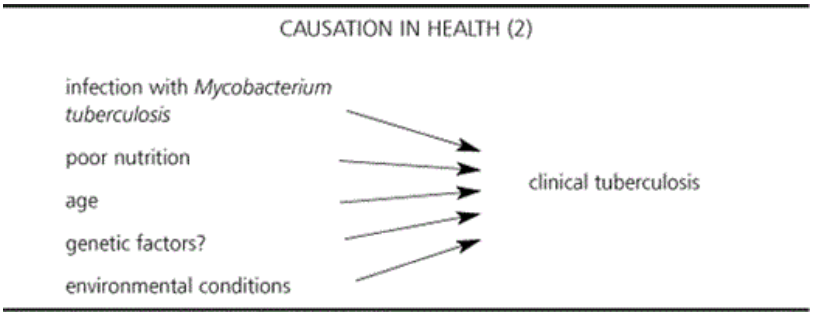


Ex. 1.2. Simple causal relationships. Causal relationships in the physical sciences are often simple, as in Boyle's law relating pressure and volume of a gas, or the effect of heat on a metal bar. The causal agent is sufficient, the time relationship is short, and replication is easy

This type of causation has a number of special properties. The chief among them is that the causal agent is
p. 7 *sufficient*, in other words the operation of the ↳ one defined causal agent invariably produces the outcome. Secondly, the time relationship between the action of the causal agent and the effect produced is very short. Thirdly, because the situation can be modelled in an experiment, it can be replicated with ease under controlled conditions.

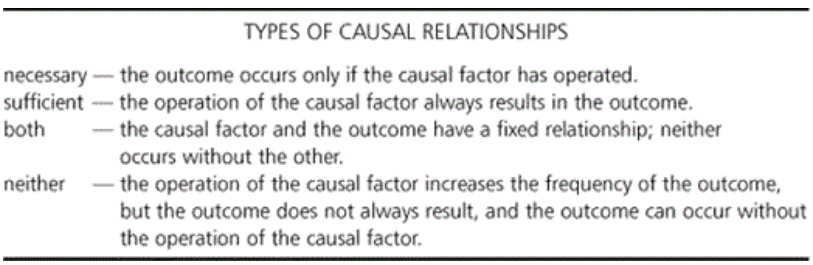
Situations that are relevant to human health and disease are rarely as simple. In human health and disease not all causal agents are sufficient. For example, the disease tuberculosis is caused by infection of the human body by the tubercle bacillus (**Ex.** 1.3). However, infection by the tubercle bacillus does not invariably lead to clinical tuberculosis. Only a small proportion of those who are infected by the bacillus develop clinical disease, and a number of other factors influence whether the disease develops, such as poor nutrition (**Ex.** 1.4). Thus a combination of tubercle bacillus infection and poor nutrition, with perhaps other factors, is required for clinical disease. We must consider both the tubercle bacillus and poor nutrition as causal factors; indeed improved nutrition was the main cause of the reduction in tuberculosis in the first half of the twentieth century.

Ex. 1.3. Simple causal relationship? An apparently simple causal relationship in medicine. But while the causal agent is necessary, it is not sufficient and the time relationship is uncertain



Ex. 1.4. Simple causal relationship? A somewhat more complete diagram of the causes of clinical tuberculosis

The tubercle bacillus in this situation is a 'necessary' cause of the disease. We can now define two categories of causal factors. A s*ufficient* causal factor, acting on its own, will always produce the outcome. A *necessary* causal
p. 8 factor is one without which the outcome cannot occur; it has acted in all instances of the ↳ outcome (**Ex.** 1.5). In the Boyle's law situation, a change in pressure was both necessary and sufficient for a change in volume, given that the other circumstances were fixed. In the metal bar example, heat was a sufficient but not a necessary cause; there are other ways of lengthening a metal bar.



Ex. 1.5. The different types of causal relationships. The last category is by far the most important

Most situations in health and disease do not fulfil the criteria for either necessary or sufficient causation. If an otherwise healthy man is admitted to hospital with multiple fractures, having been hit by a bus just outside the hospital, we would conclude that there was a causal relationship between being hit by the bus and having multiple fractures. But the relationship implies neither that the cause is sufficient nor that it is necessary. Not all people hit by buses have multiple fractures. Not all patients with multiple fractures have been hit by buses.

A frequent error of logic is to define causation in a way that describes only the limiting case of the necessary and sufficient cause. Thus the fact that Uncle Joe is alive and well at age 95, having smoked 20 cigarettes a day
p. 9 since age 10, does not ↳ show that smoking is not harmful; it shows only that smoking is not a sufficient cause of death or disability before age 95. This one observation can disprove a hypothesis of sufficient causation. But to assume, on this evidence, that no type of causation exists between smoking and disease is like concluding that the existence of elderly veterans of a war means that war does not kill people.

We must use a concept of causation that is relevant to health issues, and define a causal association in a way that is applicable to real situations. The definition of cause we will use is: *a factor is a cause of an event if its operation increases the frequency of the event* (**Ex.** 1.6).

DEFINITION OF CAUSE

cause: a factor is a cause of an event if its operation increases the frequency of the event

Ex. 1.6. The general definition of cause. Necessary and sufficient causation are merely extremes within this definition

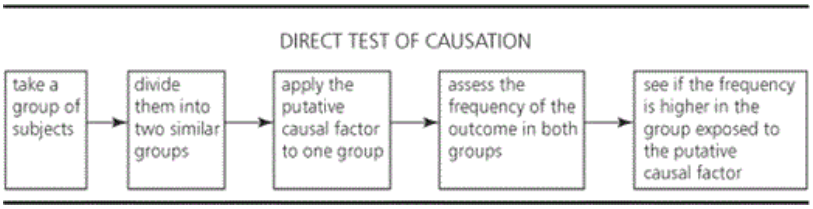
The opposite of a causal factor is, of course, a preventive factor, whose operation decreases the frequency of the outcome event. The concept of causality, and the evidence required to assess it, applies equally to preventive issues.

Returning to the man with multiple fractures, what led us to the conclusion that these were caused by having been hit by a bus? The main evidence is the immediate time relationship. Consider another situation. A woman has persistent headaches and gives a history of a head injury some months previously. Can we assume a causal relationship in this case? We cannot easily make such a judgement, because the time relation is not so clear and so many other events may have occurred. Suppose, however, we study a number of patients who have persistent headaches, and we find that most of them give a history of a head injury. We should then be justified in suspecting that the apparent connection between the two events indicates a cause and effect relationship. However, we need to be cautious about our method of recording a history of previous injury. Some people who have had an injury may not be able to remember it, and similarly many of the injuries reported might be expected as part of normal life. To go much further we need to ask whether the *frequency* with which such events have been experienced by subjects with persistent headaches is different from 'what we would expect'. To know 'what we would expect', we need to determine, by comparable methods, the frequency of such injuries in similar people who do not have persistent headaches.

So, where the time relation is not clear, and the concepts of necessary and sufficient cause do not hold, we need a *quantitative* assessment of the relationship, based on observations not on one individual but on a number of individuals. Hence the definition of causation is quantitative.

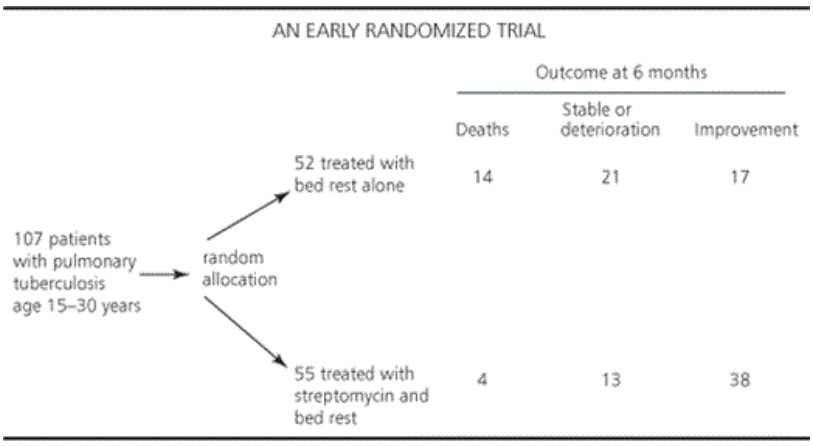
A direct test of causation: the randomized trial

A direct test of the quantitative definition of causation is given if we have two groups of individuals, who are very similar in all relevant characteristics, and apply to one group the putative causative factor. 'Relevant' characteristics are those factors, other than the one under study, which affect the frequency of the outcome event. If a causal relationship exists, the frequency of the defined outcome will be higher in the group exposed to the causal factor. A study design that uses this approach is the *randomized trial*, i.e. the assignment of the treatment for each subject is made by a random or chance procedure (**Ex.** 1.7).



Ex. 1.7. One way to test causation. When applied to therapy for patients, this is the randomized clinical trial. It is also the basic controlled experiment of biology and other sciences

Randomized trials were used first in agricultural science, where, for example, fertilizers were applied to plots of soil and identical untreated plots were used for comparison; subsequently, crop yields were measured. The causal relationship between fertilizer application and crop yield was seen in the difference in crop yield between treated and untreated plots. The application of this technique to medicine dates from the late 1940s; a trial of therapy for tuberculosis (**Ex.** 1.8) is often regarded as the first such study [1], although some other early clinical trials will be mentioned in Chapter 2. The method has been used most extensively in trials of methods of treatment, where the methods compared are generally similar, for example the choice between giving one drug and giving a different drug. The ways in which randomization contributes to the study design will be discussed in later chapters; at present it is useful to note the two main features. Because the groups receiving each of the treatments being compared are selected by a chance assignment, they are likely to be similar to each other in regard to any factors (other than the treatment) which influence the outcome; therefore differences in the outcome can be attributed to the treatment. Because the different groups are being treated and assessed at the same time, it is possible to design the study so that neither the subjects nor the investigators are aware of which treatment has been given, and this *double-blind* method will allow the observations of outcome to be made in the same way for all subjects. Since that first trial, many thousand randomized trials have been carried out.



Ex. 1.8. A trial of chemotherapy for advanced pulmonary tuberculosis, which was organized by the British Medical Research Council and which began in 1946. Often regarded as the first randomized clinical trial. From: Medical Research Council [1]

p. 11 As an aside, the tuberculosis trial arose because there was only a small amount of streptomycin in the UK immediately after the Second World War, and most of it was used for patients with advanced tuberculosis, for which the benefits were clear even without a trial, as in the absence of treatment nearly all patients would die [2]. The distinguishing mark of this trial was its use of true randomization, using random sampling numbers and sealed envelopes. The trial involved six centres, and subsequently a combined analysis of this and later trials was published, which is one of the first meta-analyses (meta-analysis will be covered in Chapter 8). There was no informed consent procedure as it was felt that if more therapy were available it would be used as the treatment of choice without specific consent, and that the consent situation for trials should not differ from routine practice.

There is no argument with the statement that the 'double-blind randomized prospective trial' is the ideal way to test a causal relationship applying to human subjects. However the procedure is not simple to carry out, not so much for scientific reasons as for practical and ethical ones. For example, to compare two types of treatment whose aim is to reduce the mortality from a chronic disease such as cancer, a large number of patients who all have a similar type of disease will have to be randomized, receive treatment, and be followed up for many years. Such a trial is obviously an expensive undertaking and will require many logistical problems to be overcome. Its results apply to the therapies used during the trial, which may be made obsolete by new therapies by the
p. 12 time the results of the trial are available. The absolute limitation of the prospective randomized ↳ trial is that it can be used only to assess interventions that are likely to be beneficial. It is clearly unethical to randomize individuals to receive a potentially harmful exposure, such as being exposed to smoking, industrial pollutants, or other toxins. Even when the comparisons are to be made between exposures that we would hope are beneficial, for example different treatments, there may well be sufficient disparity of professional opinion that many individual physicians may not judge it ethical to enter patients under their care into randomized trials.

One response to this situation is to define causation as that which can be demonstrated in a randomized trial, and to discount all other situations. This is an unrealistic stance. On this definition, we could not assume a causal relationship between the multiple fractures and the bus accident. This viewpoint that a particular scientific method determines the definition of causation is the root of the commonly made assertion that epidemiological or observational studies can show 'only associations'. This is not true. In the mid-1970s, several case–control studies (an epidemiological study design which will be explained in Chapter 2) suggested that women using hormonal therapy for menopausal symptoms had a substantially increased risk of cancer of the uterus. As hormonal therapy had been used for many years, the issue was very controversial and many experts strongly criticized these studies. Some of the criticism was that these were not double-blind randomized prospective trials. In a letter in response to a *Lancet* editorial which took this approach, it was pointed out that to ignore evidence other than that derived from randomized trials 'is to dismiss most of our

working understanding of biology, including, for example, our presumptions about the cause of that other common intrauterine tumour, pregnancy' [3].

The argument presented in this book is that the decision as to whether a certain relationship is causal must in all cases be a balanced professional judgement. We will argue that even prospective randomized trials do not 'prove' causation, because even these trials do not give certainty; there are many practical and methodological limitations of their results. It will be argued that causation can be adequately demonstrated by other scientific methods, and that the acceptance of causation on the basis of results of such methods is essential to health care practice and planning. Only a small fraction of therapeutic decisions in health care can be supported by the results of randomized trials. Very few managerial or policy decisions can be so supported. Few conclusions as to the causes and natural history of disease can be supported by such evidence. For most of our knowledge of human health in terms of therapy, aetiology, and health care planning, we must use observational studies because of the ethical and logistical impossibility of mounting randomized trials in more than a tiny
p. 13 proportion of circumstances. Therefore we must be ↳ skilful in assessing such evidence and judging whether it supports a causal relationship.

Methods of counting events: incidence and prevalence

We have seen that causation needs to be defined in terms of probabilities, and so we need to use quantitative methods. This next section reviews the terminology used in counting events in epidemiology, which may well be familiar ground to many readers.

The simplest measure of frequency is *prevalence*: the frequency of a characteristic, or the proportion of a group which has the characteristic, at one point in time. It is a simple proportion, has no units, and is measured at one point in time. In **Ex.** 1.9, there is a population of 100 000 in a community and a survey shows that at one point in time there are 800 cases of a disease present: the prevalence rate (or ratio) is 800 per 100 000. The following are real examples of prevalence rates. In a survey of 1547 people aged over 65 years in London, 563 (36 per cent) had cataract in one or both eyes; the prevalence rose from 16 per cent in men aged 65–69 to 76 per cent in women aged 85–100 [4]. This is a prevalence measure as it applies to the frequency of a condition at one point
p. 14 in time. ↳ Similarly, we could say that 23 per cent of students in a university are smokers, 25 per cent are obese, and 8 per cent have red hair. In a series of autopsies, pathologists might find that the prevalence of arteriosclerosis was 60 per cent (this in particular seems to be very frequently mislabelled as incidence in pathology texts and conversation). A slightly more difficult one is that the frequency of Down syndrome in live births is about 1.5 per 1000; this is a prevalence rate, as it is the proportion of term births which show the syndrome, and may be referred to specifically as a 'prevalence rate at birth'.

INCIDENCE, PREVALENCE, AND DURATION		
	Time point	*Time period*
	31 Dec 2001	1 Jan 2001 – 31 Dec 2002 (2 years)
Population (average or at one time point)	100 000	100 000
Number of cases at one time point	800 = prevalence prevalence rate = 800 per 100 000	
Number of new cases diagnosed over a time period		400 = incidence incidence rate = 200/100 000 per year
Number of deaths over a time period		40 = mortality mortality rate = 20/100 000 per year
Average duration of disease		$P = I \times D$ so $D = P/I$ = 800/200 = 4 years

Ex. 1.9. The relationship between prevalence, incidence, and duration of disease, assuming a steady state and a substantial population size

An *incidence rate* is the frequency of incidents, events such as deaths or new diagnoses of disease, over a defined time period; it has units of time^{-1}. For example, an annual mortality rate is the number of deaths occurring in 1 year divided by the population at risk. The annual incidence of a disease is the number of new cases occurring in a year, while the prevalence at a particular date is the number of cases existing at that date. Incidence rates are numbers of cases divided by the population size. In Ex. 1.9, 400 new cases of the disease are diagnosed over 2 years, and so the incidence rate is 200 per 100 000 per year. In Australia in 2000 there was an average population of 623 134 women aged 50–54, and there were 1562 newly diagnosed cases of breast cancer, giving an incidence rate in this age group of 250.7 per 100 000 per year. In New Zealand in 1990, there were 5914 deaths from circulatory system causes in a total population of 3 379 200 people. Therefore the mortality rate for the whole population from this cause was 175 per 100 000 per year. In a study of workers at an atomic weapons establishment, 22 552 workers were identified and the average follow-up was 18.6 years, so that the total person-years of follow-up was 22 552 × 18.6 = 419 467 person-years. There were 3115 deaths during this period, giving an average death rate over the whole time period of 7.43 deaths per 1000 person-years [5]. Therefore measurement of the incidence rate requires counting events over a period of time. The term is frequently misused in place of prevalence.

In all these examples of incidence rate, the number of events (incidents) is precisely recorded, but the denominator is an approximation, representing the average or typical number of individuals at risk, or number of person-time units at risk in a whole study. This is routinely done in analysis of vital statistics and other large bodies of data. Obviously, this type of approximation is not good enough in small sets of data. More precise analysis involves life-table methods, in which the number of individuals actually at risk at any point in time is calculated precisely, and incidence rate is based on these calculations. These methods are described in Chapter 7 (p. 264).

The *population at risk* is the appropriate denominator for many rates, although in practice the total population
p. 15 is often used. The population at risk ↳ may be defined quite precisely in detailed studies. For example, although men can get breast cancer, in general breast cancer rates are related to the female population, and in more detailed studies the population at risk of newly incidence breast cancer may be defined more accurately by excluding women who have already developed the disease. In studying an infectious disease, the population at risk may exclude immunized subjects; in a food poisoning outbreak, the population at risk from each type of food can be defined as those who consumed that food.

Cumulative incidence is the proportion of a group of subjects which experiences an event from the start to the end of a specified time period, i.e. the cumulative frequency of the event. Being a measure based on incidents it requires counting events over a period of time, but as it related to one point in time, it is a simple proportion with no units. In a study of retinopathy (damage to the retina of the eye) in subjects with diabetes, the cumulative incidence by 1 year in 3743 subjects with no retinopathy at the baseline survey and an average age of 63 was 5.3 per cent, rising to 38.1 per cent by five years [6]. The outcome from major chronic diseases in patients is often expressed as a fatality rate up to a certain point in time; for example, the cumulative incidence of death for patients diagnosed with breast cancer in Norway between 1968 and 1975 reached 40 per cent at the end of 5 years after diagnosis, and, correspondingly, the survival at that point (the 5-year survival rate) was 60 per cent. In a major study of treatment of myocardial infarction, the outcome measured was cumulative mortality over 35 days since the infarction, which was about 7 per cent; i.e. by the end of 35 days, 7 per cent of the original group had died and 93 per cent were still alive [7].

If an infectious disease, or a single-source epidemic such as food poisoning, passes through a community, at the end of the epidemic we can calculate the proportion of all people who were infected as a cumulative incidence rate over the whole time period; we might find that 20 per cent of the population was affected. This is the cumulative incidence, and is also referred to as the *attack rate*. This rate is often calculated in terms of the population at risk; for example, the attack rate of measles may be calculated specifically for those who are susceptible, excluding others who are not because of immunization.

Relationship between prevalence and incidence

Where a disease is in a stable situation in a large population (a situation of 'dynamic equilibrium'), the number of people who have a disease at one point in time (the prevalence) will depend on the incidence rate (the rate at which people develop the disease) multiplied by the average duration of the disease, ending in recovery or
p. 16 death. In fact, prevalence rate (P) equals ↳ average incidence rate (I) multiplied by average duration (D), or $P = I \times D$. Thus, if there are 200 new cases of disease each year in a community, as shown in Ex. 1.9, and the disease lasts 4 years on average, the prevalence at one point in time will be 800 cases. The steady state assumption also means that, to keep the prevalence constant, 200 cases each year must finish; as 40 deaths each year are shown in **Ex.** 1.9, this implies that 40 of the 200 cases (20 per cent) are fatal, and the other people affected are either cured or die from other causes.

Self-test questions (answers on p. 491)

Q1.1 Give a definition of a cause or causal factor.

Q1.2 What type of causation is shown in each of the following?

(a) A person falls 50 feet off a roof onto concrete and sustains a fractured leg.

(b) Exposure to the poliomyelitis virus and clinical polio.

(c) A man smokes heavily from age 14 and develops lung cancer at age 52.

(d) The association of extra chromosome 21 material with Down syndrome.

(e) A change in health service purchasing policy and a reduction in health care costs.

Q1.3 What rate, and what units, are given by:

(a) Out of 40 children in a class, 8 wear glasses.

(b) In a community of 50 000 people, there have been 100 road accidents in the last 2 years.

(c) Of 100 patients who have a certain surgical operation, 14 die in hospital.

(d) In an autopsy series, microscopic evidence of breast malignancy is found in 10 of 40 examinations.

Q1.4 Of 1000 obese subjects, 160 already have hypertension. Over 3 years, 120 more are diagnosed with hypertension.

(a) What is the initial, and the final, prevalence of hypertension?

(b) What is the cumulative incidence over 3 years in the population at risk?

(c) What is the average annual incidence rate?

Q1.5 Consider the following data on HIV seroconversion (from negative to positive) in a high-risk group of 1000 subjects. Assume that HIV ↳ positivity is permanent, and there are no entries to or losses from the group of subjects. On 1 January 2001 there are 180 positives, on 1 January 2002 there are 200 positives, and on 1 January 2003 there are 250 positives. Calculate:

p. 17

(a) the population at risk on 1 January 2001

(b) the prevalence rate at 1 January 2002

(c) the cumulative incidence over 2 years from 1 January 2001

(d) the incidence rate in 2002.

References

1. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *BMJ* 1948; **ii**: 769–782.
WorldCat

2. Yoshioka A. Use of randomisation in the Medical Research Council's clinical trial of streptomycin in pulmonary tuberculosis in the 1940s. *BMJ* 1998; **317**: 1220–1223. 10.1136/bmj.317.7167.1220
WorldCat Crossref

3. Mack TM, Pike MC. Hormone replacement therapy and endometrial carcinoma. *Lancet* 1977; **i**(8026): 1358. 10.1016/S0140-6736(77)92565-X
WorldCat Crossref

4. Reidy A, Minassian DC, Joseph J, *et al*. Prevalence of serious eye disease and visual impairment in a north London population: population based, cross sectional study. *BMJ* 1998; **316**: 1643–1646. 10.1136/bmj.316.7145.1643
WorldCat Crossref

5. Beral V, Fraser P, Carpenter L, Booth M, Brown A, Rose G. Mortality of employees of the Atomic Weapons Establishment, 1951–82. *BMJ* 1988; **297**: 757–770. 10.1136/bmj.297.6651.757
WorldCat Crossref

6. Younis N, Broadbent DM, Vora JP, Harding SP. Incidence of sight-threatening retinopathy in patients with type 2 diabetes in the Liverpool Diabetic Eye Study: a cohort study. *Lancet* 2003; **361**: 195–200. 10.1016/S0140-6736(03)12267-2
WorldCat Crossref

7. ISIS-4 (Fourth International Study of Infarct Survival Collaborative Group). ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial
p. 18 infarction. *Lancet* 1995; **345**: 669–685. ↳ 10.1016/S0140-6736(95)90865-X
WorldCat Crossref