1

Philosophy of Science and Introduction to Epidemiology

Introduction and Learning Objectives

In this chapter, we will begin by looking at different approaches to ***scientific research***, how these have arisen, and the importance of recognising that there is no single, right way to carry out investigations in the health field. Rather, we will see that different perspectives can be complementary in providing a more complete understanding of any given issue. We will then go on to explore the ***research task,*** discuss what is meant by ***epidemiology*** and ***statistics***, and look at how these two disciplines are introduced and developed in the book. The next section introduces the concept of ***rates*** for measuring the frequency of disease or characteristics we are interested in, and in particular the terms ***incidence*** and ***prevalence***. These definitions and uses of rates are fundamental ideas with which you should be familiar before we look in more detail at research methods and study design. In the final section, we will look at key concepts in disease prevention, including the commonly used terms ***primary***, ***secondary***, ***and tertiary*** prevention.

The reason for starting with a brief exploration of the nature of scientific methods is to see how historical and social factors have influenced the biomedical and social research traditions that we take for granted today. This will help you understand your own perceptions of, and assumptions about, health research, based on the knowledge and experience you have gained to date. It will also help you understand the scientific approach being taken in this book and how this both complements and differs from that developed in books and courses on qualitative research methods – as and when you may choose to study these. Being able to draw on a range of research traditions and their associated methods is especially important for the discipline of public health, but it is also important for many other aspects of health and health care.

Learning Objectives

By the end of this chapter, you should be able to do the following:

- Briefly describe the key differences between the main approaches to research that are used in the health field.
- Describe what is meant by epidemiology, and list the main uses to which epidemiological methods and thought can be put.
- Describe what is meant by statistics, and list the main uses to which statistical methods and thought can be put.
- Define and calculate rates, prevalence, and incidence, and give examples of their use.
- Define primary, secondary, and tertiary prevention and give examples of each.

Quantitative Methods for Health Research: A Practical Interactive Guide to Epidemiology and Statistics, Second Edition.
Nigel Bruce, Daniel Pope, and Debbi Stanistreet.

Companion Website: www.wiley.com/go/bruce/quantitative-health-research

1.1 Approaches to Scientific Research

1.1.1 History and Nature of Scientific Research

Scientific research in health has a long history going back at least to the classical period. There are threads of continuity, as well as new developments in thinking and techniques, that can be traced from the ancient Greeks and through the fall of the Roman Empire, the Dark Ages, and the Renaissance to the present time. At each stage, science has influenced, and has been influenced by, the culture and philosophy of the time. Modern scientific methods reflect these varied historical and social influences. So it is useful to begin this brief exploration of scientific health research by reflecting on our own perceptions of science and how our own views of the world fit with the various ways research can be approached. As you read this chapter you might like to think about the following questions:

- What do you understand by the terms ***science*** and ***scientific research***, especially in relation to health?
- How has your understanding of research developed?
- What type of research philosophy best fits your view of the world and the health issues you are most interested in?

Thinking about the answers to these questions will help you understand what we are trying to achieve in this section and how this can best support the research interests that you have and are likely to develop in the years to come. The history and philosophy of science is of course a whole subject in its own right, and this is of necessity a very brief introduction.

Scientific Reasoning and Epidemiology

Health research involves many different scientific disciplines, many of which you will be familiar with from previous training and experience. Here we are focusing principally on epidemiology, which is concerned with the study of the distribution and determinants of disease within and between populations. In epidemiology, as we shall see, there is an emphasis on ***empiricism,*** that is, the study of observable phenomena by scientific methods, detailed observation, and accurate measurement. The scientific approach to epidemiological investigation has been described as

- **Systematic** – There is an agreed system for performing observations and measurement.
- **Rigorous** – The agreed system is followed exactly as prescribed.
- **Reproducible** – All the techniques, apparatus, and materials used in making the observations and measurements are written down in enough detail to allow another scientist to reproduce the same process.
- **Repeatable** – Scientists often repeat their own observations and measurements several times in order to increase the reliability of the data. If similar results are obtained each time, the researcher can be more confident the phenomena have been accurately recorded.

These are characteristics of most epidemiological study designs and are an important part of the planning and implementation of the research. However, this approach is often taken for granted by many investigators in the health field (including epidemiologists) as the only way to conduct research. Later we will look at some of the criticisms of this approach to scientific research, but first we need to look in more detail at the reasoning behind this perspective.

Positivism

The view of science and knowledge known as ***positivism*** is the dominant philosophy underlying contemporary epidemiology. The evolution of positivism has been extensively documented elsewhere (see Guba and Lincoln, 1994; Halfpenny, 1982; and Feigl, 1969), its early development being attributed mainly to August Comte during the early 19th century. However, its roots can be traced back to the 17th century, to a time when scientists stopped relying on religion, conjecture, and faith to explain phenomena, and instead began to use reason and rational thought. This period saw the emergence of the view that it is only by using scientific thinking and practices that we can reveal the truth about the world.

Positivism assumes a stable observable reality that can be measured and observed. So, for positivists, scientific knowledge is proven knowledge, and theories are therefore derived in a systematic, rigorous way from observation and experiment. This approach to studying human life is the same approach that scientists take to study the natural world. Human beings are believed by positivists to exist in causal relationships that can be empirically observed, tested, and measured (Bilton *et al*., 2002) and to behave in accordance with various laws. Because this reality exists whether we look for it or not, it is the role of scientists to reveal its existence but not to attempt to understand the inner meanings of these laws or express personal opinions about these laws. One of the primary characteristics of a positivist approach is that the researcher takes an objective distance from the phenomena so that the description of the investigation can be detached and undistorted by emotion or personal bias (Davey, 1994). This means that within epidemiology, various study designs and techniques have been developed to increase objectivity; you will learn more about these in later chapters.

More recently, some of the earlier tenets of positivism have been challenged through the work of Karl Popper and other scholars such as Bronowski (Bronowski, 1956; Popper, 1959), and as a result, a post-positivistic approach has emerged since the mid-20th century or so, and this approach now underpins much contemporary empirical research activity (Philips, 1990). Post-positivism still advocates that there is an objective reality, but it suggests that reality can only be measured imperfectly due to the limitations of the scientific approach. It also asserts a realist perspective, stating that there are phenomena that can't be observed but nevertheless do exist, so science is not limited to only those phenomena that can be measured. A more-detailed discussion of post-positivism is outside of the scope of this book. For the purposes of this chapter, when we refer to positivism, this can be assumed to refer to both positivist and post-positivist approaches within epidemiology.

A second aspect of scientific thinking, which also evolved over this period, derives from the work of Thomas Kuhn (1922–1996), who challenged the concept of absolute evidence. In *The Structure of Scientific Revolutions*, (Kuhn, 1970), Kuhn argued that one scientific paradigm – one 'conceptual worldview' – may be dominant at a particular period in history. Over time, however, this paradigm is challenged, and eventually it is replaced by another view (paradigm), which then becomes accepted as the most important and influential. He termed these revolutions in science 'paradigm shifts'. Although questioned by other writers, this perspective suggests that scientific methods we may take for granted as being the only or best way to investigate health and disease are to an extent the product of historical and social factors, and they can be expected to evolve – and maybe change substantively – over time.

Induction and Deduction

There are two main forms of scientific reasoning: ***induction*** and ***deduction***. Both have been important in the development of scientific knowledge, and it is useful to appreciate the difference between the two in order to understand the approach taken in epidemiology.

Induction

With inductive reasoning, researchers make repeated observations and use this evidence to generate theories to explain what they have observed. For example, if a researcher made a number of observations in different settings of women cooking dinner for their partners, they might then inductively derive a general theory:

> All women cook dinner for their partners.

Deduction

Deduction works in the opposite way to induction, starting with a theory (known as an ***hypothesis***) and then testing it by observation. Thus, a very important part of deductive reasoning is the formulation of the hypothesis – that is, the provisional assumption researchers make about the population or phenomena they wish to study before starting with observations. A good hypothesis must enable the researcher to test it through a series of ***empirical observations.*** So, in deductive reasoning, the hypothesis would be

> All women will cook dinner for their partners.

Observations would then be made in order to test the validity of this statement. This would allow researchers to check the consistency of the hypothesis against their observations, and if necessary, the hypothesis can be discarded or refined to accommodate the observed data. So, if they found even one woman not cooking for her partner, the hypothesis would have to be re-examined and modified. This characterises the approach taken in epidemiology and by positivists generally.

Karl Popper (1902–1994) argued that hypotheses can never be proved true for all time, and scientists should aim to refute their own hypotheses even if this goes against what they believe (Popper, 1959). He called this the ***hypothetico-deductive method***, and in practice this means that an hypothesis should be capable of being falsified and then modified. Thus, to be able to claim the hypothesis is true would mean that all routes of investigation have been carried out. In practice, this is impossible, so research following this method does not set out with the intention of proving that an hypothesis is true. In due course we will see how important this approach is for epidemiology and in the statistical methods used for testing hypotheses.

Alternative Approaches to Research

It is important to be aware that positivism is only one of many different approaches to scientific research. Many social scientists, for example, believe that these approaches are not relevant for the study of human behaviour. From this perspective, they believe that human beings do not act in accordance with observable rules or laws. This makes humans different from phenomena in the natural world, and so they need to be studied in a different way. Positivism (and post-positivism) have also been criticised because they cannot explain how people interpret or make sense of the world. As Green and Thorogood (2004, p. 12) argue,

> Unlike atoms (or plants or planets), human beings make sense of their place in the world, have views on researchers who are studying them, and behave in ways that are not determined in law-like ways. They are complex, unpredictable, and reflect on their behaviour. Therefore, the methods and aims of the natural sciences are unlikely to be useful for studying people and social behaviour: instead of explaining people and society, research should aim to understand human behaviour.

Many social scientists therefore hold different beliefs about how we should carry out research into human behaviour. Consequently, they are more likely to take an ***inductive*** approach to research because they argue that they do not want to make assumptions about the social world until they have observed it in and for itself. They therefore do not want to formulate hypotheses because they believe these are inappropriate for making sense of human action. Rather, they believe that human action cannot be explained but must be understood.

Whereas positivists would be concerned mainly with observing patterns of human behaviour, other researchers principally wish to understand that behaviour. This latter group requires a different starting point that will encompass their view of the world, or different ***theoretical positions*** to make sense of the world. It turns out that there are many different positions that can be adopted, and while we cannot go into them all here, we briefly consider one of the most important of these, known as an ***interpretative*** approach.

An interpretative approach assumes an interest in the meanings underpinning human action, and the role of the researcher is therefore to unearth that meaning. The researcher would not look to measure the reality of the world but would seek to understand how people interpret the world around them (Green and Thorogood, 2004).

Let's look at an example of positivist and interpretivist approaches in respect of a common health problem with multiple physiological, social, and behavioural aspects, namely, asthma. A ***positivist*** approach to researching this condition may be to obtain a series of objective measurements of symptoms and lung function using a standard procedure on a particular sample of people over a specified period of time. An ***interpretative*** approach might involve talking in-depth to a small number of participants with asthma to try to understand how they view the impact of their symptoms on their lives. Obviously, in order to do this, these two types of approaches require the use of different research methods. Those planning interpretative research would use ***qualitative methods*** (e.g. interviews, focus groups, and ethnographic methods), whereas positivists (e.g. epidemiologists) would choose ***quantitative methods*** (e.g. surveys and cohort studies involving lung-function measurements and highly structured questionnaires). These two different approaches would draw on different ***research paradigms*** and would therefore produce different types of findings.

Those drawing on an interpretative perspective would also differ from positivists in respect of the view that researchers can have an objective, unimpaired, and unprejudiced stance in the research that allows them to make value-free statements. Interpretative research accepts that researchers are human beings and therefore cannot stand objectively apart from the research. In a sense, they are part of the research process, as their presence can influence the nature and outcome of the investigation.

With the asthma example, we can see how complementary the findings of these two different approaches to research could be. The positivist approach can help determine whether a new medication provides any benefit in terms of control of symptoms or lung function, for example. On the other hand, the interpretative approach can help us understand why the activities of some people may be more affected by their condition than others, for example.

Researchers working with a post-positivist framework have to some extent narrowed the divide between quantitative and qualitative approaches to research, since post-positivism does not reject approaches that focus on the meanings people give to their actions as seen in interpretative approaches to research. Mixed-methods approaches (using both qualitative and quantitative methods) to research can therefore help build up a more-complete picture of effective health care and support that allows a better understanding of many aspects of a person's life and health experience. Further discussion of mixed methods is outside of the scope of this book; for a useful guide, see Cresswell and Piano Clark (2011).

JWST868-c01 JWST868-Bruce October 25, 2017 9:56 Printer Name: Trim: 254mm × 178mm

Exercise for Reflection

1. Make brief notes on the type of scientific knowledge and research with which you are most familiar.
2. Is this predominantly positivistic (hypothetico-deductive) or interpretative in nature, or is it more of a mixture?

There are no answers provided for this exercise, as it is intended for personal reflection.

1.1.2 What is Epidemiology?

The term *epidemiology* is derived from the following three Greek words:

Epi **– among**
Demos **– the people**
Logos **– study of**

We can translate this in more modern terms into ***the study of the distribution and determinants of disease frequency in human populations***. The following exercise will help you to think about the uses to which the discipline of epidemiology is put.

Self-Assessment Exercise 1.1.1

Make a list of some of the ***applications*** of epidemiological methods and thought that you can think of. In answering this, avoid listing types of epidemiological study that you may already know. Try instead to think in general terms about the practical applications of these methods.

Answers in Section 1.5

This exercise shows the very wide application of epidemiological methods and thought. It is useful to distinguish between two broad functions of epidemiology, one very practical, the other more philosophical:

- The range of epidemiological research methods provides a toolbox for obtaining the best scientific information in a given situation (assuming, that is, you have established that a positivist approach is most appropriate for the topic under study!).
- Epidemiology helps us use knowledge about the population determinants of health and disease to inform the full range of investigative work, from the choice of research methods, through analysis and interpretation, to the application of findings to policy. With experience, this becomes a way of thinking about health issues over and above the mere application of good methodology.

You will find that your understanding of this second point grows as you learn about epidemiological methods and their application. This is because epidemiology provides the means of describing the characteristics of populations, comparing them, and analysing and interpreting the differences, as well as the many social, economic, environmental, behavioural, ecological, and genetic factors that determine those differences.

1.1.3 What are Statistics?

A statistic is a numerical fact. Your height and weight and the average daily rainfall in Liverpool are examples of statistics. The academic discipline of statistics is concerned with the collection, presentation, analysis, and interpretation of numerical information (also called ***quantitative*** information).

Statistics are Everywhere!

We are surrounded by, and constantly bombarded with, information from many sources: the cereal box, unemployment figures, football results, opinion polls, and articles in scientific journals. The science of statistics allows us to make sense of this information and is thus a fundamental tool for investigation in many disciplines, including health, education, economics, agriculture, and politics, to name but a few. The next exercise encourages you to explore how statistics are used in everyday life.

Self-Assessment Exercise 1.1.2

Look at some general information sources such as newspapers or websites and find up to five items in which statistics are used. List the ways numerical information is presented.

Examples in Section 1.5

The scientific term for pieces of information is ***data***. The singular is ***datum***, meaning a single piece of information, such as, for example, one person's weight. A set of data may consist of many items, such as the heights, weights, blood pressures, smoking habits, and exercise levels of several hundred people. In its raw state, this mass of figures tells us little. There are two ways we use statistics to help us interpret data:

- To ***describe*** the group about which the data have been collected. This may be a group of people or a group of hospitals or a group of laboratory specimens. We describe the group by summarising the information into a few meaningful numbers and pictures. We discuss this further in Chapter 2.
- To ***infer*** something about the population of which the group we are studying is a part. We often want to know something about a population, such as everyone older than 65 years in Liverpool, but, practically, can only collect information about a subset of that population. This subset is called a sample, and it is explored in Chapter 3 on surveys. With inference, we want to know what generalisations to the population can be made from the sample and with what degree of certainty.

Self-Assessment Exercise 1.1.3

Can you find one example of ***description*** and one example of ***inference*** in your newspaper or Web search? If you have found an example of making an inference, to which population does it apply?

Examples in Section 1.5

1.1.4 Approach to Learning

We will explore the use and interpretation of statistical techniques through a number of published studies. Whether or not you go on to carry out research and use statistical methods yourself, you are certain to be a consumer of statistics through published research. We will emphasise both the use of appropriate techniques and the critical interpretation of published results. You will also be learning about epidemiology and statistics in an integrated way. This approach recognises that the two disciplines embody many closely related concepts and techniques. There are also certain very distinct qualities, which you will find are emphasised through the more theoretical discussion relating to one or another discipline. Your learning of these research methods will be based primarily on practical examples of data and published studies in order to help you to see how epidemiology and statistics are used in practice, and not just in theoretical or ideal circumstances.

Summary

- There is no single, right philosophy of health research. The approach taken is determined primarily by the nature of the problem being investigated as well as by a range of other factors, including historical and social influences.
- As an individual, your education, training, and experience strongly influence the scientific paradigm with which you are familiar and comfortable.
- A variety of approaches to, and methods for, research are both appropriate and necessary in the health field.
- Epidemiology provides us with a range of research tools, which can be used to obtain the information required for preventing health problems, providing services, and evaluating health care. One of the most important contributions of epidemiology is the insight gained about the factors that determine the health of populations.
- Statistics is concerned with the collection, presentation, analysis, and interpretation of numerical information. We may use statistical techniques to describe a group (of people, hospitals, etc.) and to make inferences about the population to which the group belongs.

1.2 Formulating a Research Question

1.2.1 Importance of a Well-Defined Research Question

This is arguably the most important section of the whole book. The reason for our saying this is that the methods we use, and ultimately the results that we obtain, must be determined by the question we are seeking to answer.

So how do we go about formulating that question? This does not (usually) happen instantaneously, and there is a good deal of debate about how the question ought to be formulated and how it is formulated in practice. For example, Karl Popper argued that research ideas can come from all kinds of sources. But the idea is not enough on its own, and it usually needs working on before it is a clearly formulated ***research question***. Here are some of the factors we have to take into account in fashioning a clear research question:

- What does all the other work on the topic tell us about the state of knowledge, and what aspects need to be addressed next? Our idea might actually arise from a review such as this and therefore be already fairly well defined, but more often than not, the idea or need arises before we have had a chance to fully evaluate the existing body of knowledge. This also depends on how far we already are into researching a particular subject.

- Different types of problem and topic areas demand, and/or have been traditionally associated with, different research traditions. Does our idea require a positivist approach with an hypothesis that can be falsified? If so, the research question needs to be phrased in a way that allows this. Alternatively, we might be trying to understand how a certain group of people view a disease and the services provided for them. This question must also be precisely defined, but not in the same way: There is nothing here to be falsified; rather, we wish to gain as full an understanding as possible of people's experience and perspectives to guide the development of services.
- If our idea is ambitious, the necessary research may be too demanding, expensive, or complex to answer the question(s) in one go. Perhaps it needs to be done in separate studies or in stages.

In practice, defining the research question does not usually happen cleanly and quickly but is a process that gradually results in a more and more sharply defined question as the existing knowledge, research options, and other practical considerations are explored and debated. There may appear to be exceptions to this – for instance, a trial of a new drug. On the face of it, the question seems simple enough: The drug is now available, so is it better than existing alternatives or not? However, as we will discover later, the context in which the drug would be used can raise a lot of issues that play a part in defining the research question.

Although knowledge of appropriate research methods is important in helping you to formulate a clear research question, it is nevertheless useful to start the ***process*** of developing your awareness of, and skills in, this all-important aspect of research. The following exercise provides an opportunity for you to try this.

Self-Assessment Exercise 1.2.1

A Research Idea …

Your work in an urban area involves aspects of the care and management of people with asthma. You are well aware of the current concern about the effect of air pollution on asthma and the view that whereas pollution (e.g. ozone, nitrogen oxides) almost certainly exacerbates asthma, this may not be the cause of the underlying asthmatic tendency.

In recent years, you have noticed that asthmatics (especially children) living in the poorest parts of the city seem to suffer more severe and frequent asthma episodes than those living in better-off parts.

Although you recognise that pollution from traffic and industry might be worse in the poorer areas, you have been wondering whether other factors, such as diet (e.g. highly processed foods) or housing conditions such as dampness and associated moulds, might be the real cause of the difference.

You have reviewed the literature on this topic and have found a few studies that are somewhat conflicting and that do not seem to have distinguished very well among the pollution, diet, and housing factors you are interested in.

Have a go at converting the idea described above into a well-formulated research question appropriate for epidemiological enquiry. Note that there is no single, right research question here. You do not need to describe the study methods you might go on to use.

Specimen Answer in Section 1.5

1.2.2 Development of Research Ideas

We have seen that research is a process that evolves over a period of time. It is influenced by many factors, including other work in the field, the prevalent scientific view, political factors, finance, and so on. Research is not a socially isolated activity with a discrete (inspirational) beginning, (perfect) middle, and (always happy) ending (Figure 1.2.1).

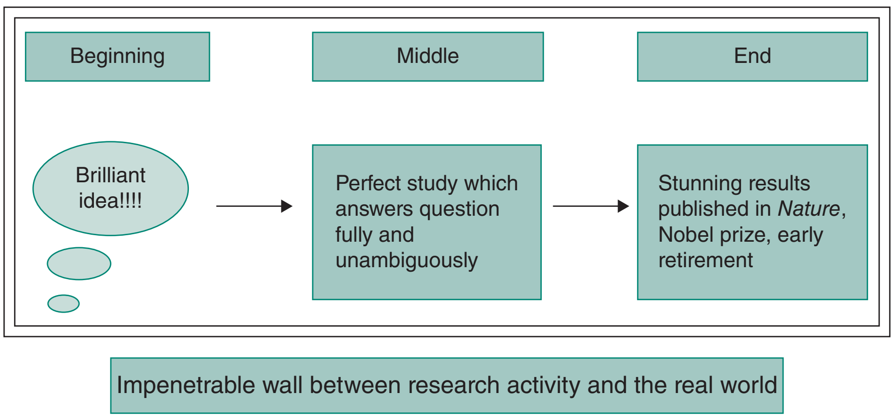


Figure 1.2.1 Research fantasy time.

A more realistic way to describe the process of research development is cyclical, as illustrated in Figure 1.2.2. A well-defined and realistic question, which is (as we have seen) influenced by many factors, leads to a study that, it is hoped, provides much of the information required.

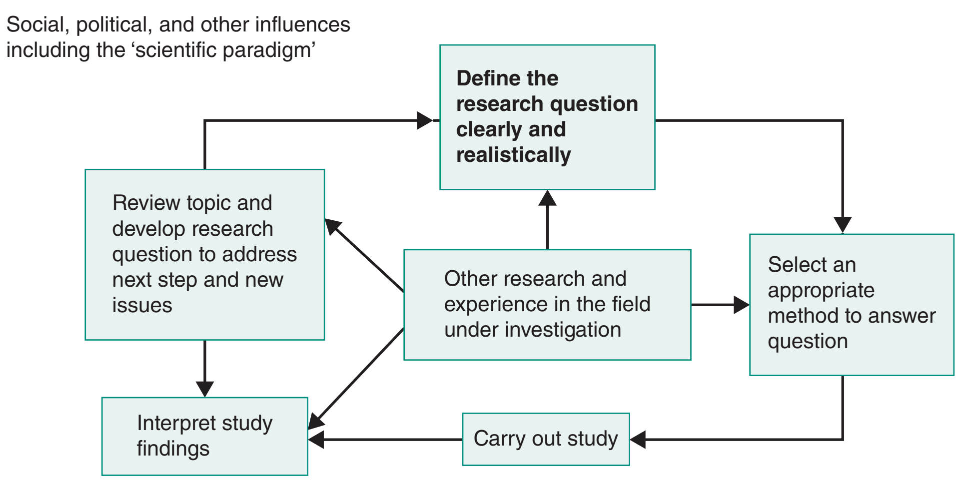


Figure 1.2.2 Research is a process that can usefully be thought of as being cyclical in nature and subject to many influences from both inside and outside the scientific community.

These findings, together with developments in the scientific field as well as social, political, or other significant influences, will lead to further development of the research question.

Summary

- Defining a clear research question is a fundamental step in research.
- Well-defined research questions do not (usually) appear instantly (although the underlying idea might arise in a moment of inspiration)!
- Research is a process, which can usefully be thought of as cyclical, albeit subject to many external influences along the way.

1.3 Rates: Incidence and Prevalence

1.3.1 Why Do We Need Rates?

Incidence and prevalence are terms you will have encountered commonly in everyday language as well as in research. In quantitative health research, they have clear and specific definitions, and they represent fundamental concepts of how the frequencies of diseases or characteristics among groups of people are described. Why then are they so important?

One way to approach this question is by considering the problem that arises when we try to interpret a change in the number of events (which could be deaths, hospital admissions, etc.) occurring during, say, a period of one year in a given setting. Exercise 1.3.1 is an example of this type of problem and is concerned with an increase in numbers of hospital admissions.

Self-Assessment Exercise 1.3.1

Over a period of 12 months, the accident and emergency department of a city hospital noted that the number of acute medical admissions for people over 65 had increased by 30 per cent. In the previous 5 years, there had been a steady increase of only about 5 per cent per year.

1. List the possible reasons for the 30 per cent increase in hospital accident and emergency admissions.
2. What other information could help us to interpret the reasons for this sudden increase in admissions?

Answers in Section 1.5

In this exercise we have seen the importance of interpreting changes in numbers of events in the light of knowledge about the ***population*** from which those events arose. This is why we need ***rates***. A rate has a ***numerator*** and a ***denominator*** and must be determined over a specified ***period of time***. It can be defined as follows:

"Numerator"

$$\text{RATE} = \frac{\textbf{Number of events arising from defined population in a given period}}{\textbf{Number in defined population, in same period}}$$

"Denominator"

1.3.2 Measures of Disease Frequency

We can view rates as measures of the ***frequency*** of disease or of characteristics in the population. Two of the most important ways of presenting this information are provided by ***prevalence*** and ***incidence*** rates.

1.3.3 Prevalence Rate

The prevalence rate tells us how many cases of a disease (or people with a characteristic, such as smoking) there are in a given population at a specified time. The ***numerator*** is the number of cases, and the ***denominator*** is the population we are interested in.

$$\boxed{\textbf{\textit{Prevalence}} = \frac{\text{Number of cases at a given time}}{\text{Number in population at that time}}}$$

This can be expressed as a percentage, or per 1,000 population, or per 10,000, etc., as convenient. The following are therefore examples of prevalence:

- In a local government area with a population of 400,000, there are 100,000 smokers. The prevalence is therefore 25 per cent, or 250 per 1,000.
- In the same area, there are known to be 5,000 people with diagnosed schizophrenia. The prevalence is therefore 1.25 per cent, or 12.5 per 1,000.

Note that these two examples represent snapshots of the situation at a given time. We do not have to ask about people starting or giving up smoking or about people becoming ill with (or recovering from) schizophrenia. It is a matter of asking, 'In this population, how many are there now?' This snapshot approach to measuring prevalence is known as ***point prevalence***, since it refers to one point in time, and it is the usual way the term *prevalence* is used. If, on the other hand, we assess prevalence over a period of time, it is necessary to think about cases that exist at the start of the period and new cases that develop during the period. This measure is known as ***period prevalence***, and this, together with point prevalence, is illustrated in Figure 1.3.1 and Exercise 1.3.2.

Point prevalence (Figure 1.3.1a) is assessed at one point in time (time A), whereas period prevalence (Figure 1.3.1b) is assessed over a period (period B). The horizontal bars represent patients becoming ill and recovering after varying periods of time – the start and end of one episode is marked in Figure 1.3.1a. Period prevalence includes everyone who has experienced the disease at some time during this period.

Self-Assessment Exercise 1.3.2

1. In the examples in Figure 1.3.1, calculate the point prevalence and period prevalence.
2. Why do the point prevalence and period prevalence differ?

Answers in Section 1.5

1.3.4 Incidence Rate

Whereas the ***prevalence*** rate gives us a measure of how many cases there are in the population at a given time (or period when period prevalence is used), the ***incidence rate*** tells us the rate at which ***new*** cases are appearing in the population over a specified time period. The time period

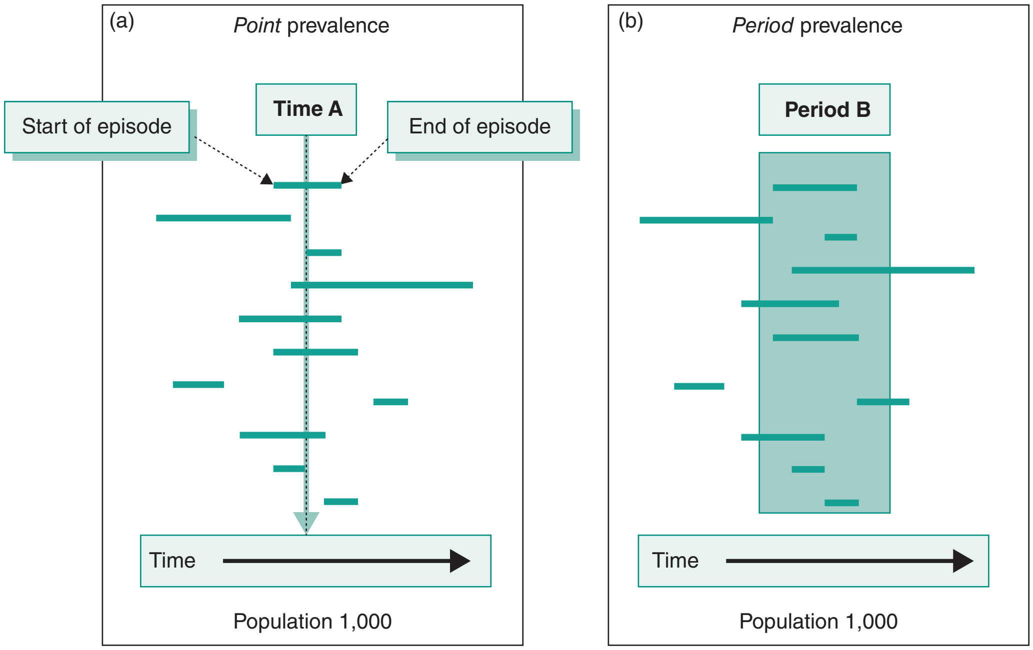


Figure 1.3.1 Period and point prevalence.

must always be specified for an incidence rate. To determine the incidence rate, we need to know the number of new cases appearing over a specified period of time and the number of people in the population who could become cases over that same period of time (the at-risk population). This is called the ***cumulative incidence rate*** and is calculated as follows:

Incidence rate (also termed 'cumulative incidence rate')

$$\text{Incidence} = \frac{\text{Number of new cases arising from a defined population in a specified time period}}{\text{Number in defined at-risk population over the same time period}}$$

This rate can be expressed per 1,000 per year (or other convenient time period) or per 10,000 per year, etc., as appropriate, over the specified time period. Thus, if there were 20 new cases of a disease in an at-risk population of 2,500 over a period of one year, the cumulative incidence rate, expressed per 1,000 per year, is as follows:

$$\text{Incidence} = \frac{20}{2{,}500} \times 1{,}000$$

This gives a rate of 8 cases per 1,000 per year. The denominator – the defined population – must be exclusively the population that could become cases (termed ***at risk***). For example, if we are considering the rate of hysterectomy (removal of the uterus) among UK women aged 50 years and older, the denominator population must be women 50 and older but excluding those who have had a hysterectomy.

Quite often, however, when a large group of the population is being studied – for example, all men in a city such as Liverpool – the exact number at risk throughout the year in this defined population will not be readily available, and an estimate has to be made. The corresponding midyear population estimate is often used as the denominator population, since it is usually available from published official statistics.

Using a source of information on the population such as this means that existing cases who are not at risk of becoming new cases of the disease (because they are already affected) will be included in the denominator of the equation (unless data on the numbers of cases of the disease in question are available to allow these to be removed). Where the number of existing cases is not known, the resulting incidence measure will not be much affected so long as the number of cases is relatively small in comparison to the total population, but you should be aware that the true incidence rate of the disease will be slightly underestimated. The following exercise will help consolidate your understanding of the cumulative incidence rate.

Self-Assessment Exercise 1.3.3

1. In Figure 1.3.1b, how many new cases arose during period B?
2. Surveillance data for episodes of food poisoning in a given population showed that among children aged 0–14 years, there had been 78 new cases among boys and 76 new cases among girls over a period of 1 year. The population breakdown for the area is as shown in the following table.

Age group (years)	Female	Male
0–14	37,100	41,000
15–24	59,610	60,100
25–44	62,050	57,300
45–64	42,450	39,610
65+	28,790	21,990
Total	230,000	220,000

Calculate the annual cumulative incidence rates per 10,000 for boys aged 0–14 years and for girls aged 0–14 years. Comment on what you find.

Answers in Section 1.5

Person-Time

The cumulative incidence rate assumes that the entire population at risk at the beginning of the study has been followed up for the same amount of time. It therefore measures the proportion of unaffected individuals who, on average, will contract the disease over the specified time period. However, in some study designs, such as cohort studies (described in Chapter 5), people may be entered into the study at different times and then be followed up to a specific end-of-study date. In addition, some might withdraw from the study or might die before the end of the study. Study participants therefore have differing lengths of follow-up.

To account for these varying times of follow-up, a denominator measure known as ***person-time*** is used. This is defined as the sum of each individual's time at risk while remaining free of disease. When *person-time* is used, incidence is calculated slightly differently and is known as the ***incidence density,*** which is the average person-time incidence rate. Note, we do not include 'rate' in the term 'incidence density', as this is implied. Incidence density is calculated as follows:

Incidence density

$$\text{Incidence} = \frac{\text{Number of new cases arising from a defined population}}{\text{Total at risk person–time of observation}}$$

Since person-time can be counted in various units such as days, months, or years, it is important to specify the time units used. For example, if six new cases of a disease are observed over a period of 30 person-years, then the incidence would be $6/30 = 0.2$ per person-year or, equivalently, 20 per 100 person-years or 200 per 1,000 person-years. If people are lost to follow-up or withdraw from the study prematurely, their time at risk is taken to be the time they were under observation in the study. This next exercise will help with understanding incidence density.

Self-Assessment Exercise 1.3.4

Time of follow-up in study, or until disease develops, for 30 subjects:

Subject number	Years of follow-up	Disease (Y or N)	Subject number	Years of follow-up	Disease (Y or N)
1	19.6	N	16	0.6	Y
2	10.8	Y	17	2.1	Y
3	14.1	Y	18	0.8	Y
4	3.5	Y	19	8.9	N
5	4.8	N	20	11.6	Y
6	4.6	Y	21	1.3	Y
7	12.2	N	22	3.4	N
8	14.0	Y	23	15.3	N
9	3.8	Y	24	8.5	Y
10	12.6	N	25	21.5	Y
11	12.8	Y	26	8.3	N
12	12.1	Y	27	0.4	Y
13	4.7	Y	28	36.5	N
14	3.2	N	29	1.1	Y
15	7.3	Y	30	1.5	Y

1. Assuming that all subjects in the table enter the study at the same time and are followed up until they leave the study (end of follow-up) or develop the disease, find the total observation time for the 30 subjects and estimate the incidence density. Give your answer per 1,000 (10^3) person-years.
2. It is more usual for follow-up studies to be of limited duration, where not all the subjects will develop the disease during the study period. Calculate the incidence density for the same 30 subjects if they were observed for only the first 5 years, and compare this with the rate obtained in question 1.

Answers in Section 1.5

1.3.5 Relationship Between Incidence, Duration, and Prevalence

There is an important relationship between the incidence rate, illness duration, and prevalence rate. Consider the following two examples:

- Urinary tract infections among women are seen very commonly in general practice, reflecting the fact that the incidence rate is high. The duration of these infections (with treatment) is

usually quite short (a few days), so at any one time there are not as many women with an infection as one might imagine given the high incidence. Thus, the (point) prevalence is not particularly high.
- Schizophrenia is a chronic psychiatric illness from which the majority of sufferers do not recover. Although the incidence is quite low, it is such a long-lasting condition that at any one time the prevalence is relatively high, between 0.5 per cent and 1 per cent, or 5 to 10 per 1,000 in the UK.

Thus, for any given incidence, a condition with a longer duration will have a higher prevalence. Mathematically, we can say that the prevalence rate is proportional to the product of the incidence rate and the average duration of the disease. In particular, when the prevalence rate is low (less than about 10 per cent), the relationship can be expressed as follows:

$$\boxed{\text{Prevalence} = \text{incidence} \times \text{duration}}$$

This formula holds so long as the units concerned are consistent, the prevalence rate is low, and the duration is constant (or an average can be taken). Thus, if the incidence rate is 15 per 1,000 per year, and the duration is, on average, 26 weeks (0.5 years), then the prevalence rate will be (15×0.5) per $1{,}000 = 7.5$ per 1,000.

Summary

- Rates are a very important concept in epidemiology, and they allow comparison of information on health and disease from different populations.
- A rate requires a numerator (cases) and a denominator (population), each relating to the same specified time period.
- The prevalence rate is the number of cases in a defined population at a particular point in time (point prevalence) or during a specified period (period prevalence).
- The incidence rate is the number of new cases that occur during a specified time period in a defined at-risk population (cumulative incidence).
- Incidence density is a more precise measure of incidence and uses person-time of observation as the denominator.
- Without using rates, comparison of numbers of cases in different populations may be very misleading.
- The relationship between incidence and prevalence is determined by the duration of the condition under consideration.

1.4 Concepts of Prevention

1.4.1 Introduction

In this section, we look at the ways we can describe ***approaches to prevention***. This is a well-established framework that provides important background to many of the studies we will examine as we learn about research methods, as well as for services such as screening.

The following examples illustrate three different approaches to disease prevention. Please read through these, and complete Exercise 1.4.1. We will then look at the formal definitions of these approaches.

Example 1: Road Accidents Among Children

In 2012, accidents were the most common cause of death among male children and young adults aged 5–19 years in the UK, and the majority of these accidents occurred on the roads. Lower speed limits, linked to stricter enforcement, offer one way of reducing the number of these deaths arising from road accidents.

Example 2: Breast Cancer

Breast cancer is one of the most common cancers among women. Despite this, we know little for certain about the causes beyond genetic, hormonal, and some dietary factors. For a number of years, mammography, a radiographic (X-ray) examination of the breast, has been routinely offered to women 50–64 years of age. Abnormalities suggestive of cancer are investigated by biopsy (removal of a small piece of tissue for microscopic examination), and if the biopsy is positive, the cancer is treated.

Example 3: Diabetes and the Prevention of Foot Problems

Diabetes, a disorder of blood glucose (blood sugar) metabolism, is generally a progressive condition. The actual underlying problem does not usually resolve, and control of blood glucose has to be achieved through attention to diet, and usually also with medication, which may be in tablet form or as injected insulin. Associated with this disordered glucose metabolism are a range of chronic degenerative problems, including atherosclerosis (which leads to heart attacks and to poor blood supply to the lower legs and feet), loss of sensation in the feet due to nerve damage, and eye problems. Many of these degenerative processes can be slowed down, and associated problems prevented, by careful management of the diabetes. One important example is care of the foot when blood supply and nerves are affected. This involves educating the diabetic patient about the problem and about how to care for the foot, and providing the necessary treatment and support.

Self-Assessment Exercise 1.4.1

For each of the above examples, describe in everyday language (that is, avoiding technical terms and jargon) how prevention is being achieved. In answering this question, think about the way the prevention measure acts on the development and progression of the disease or health problem concerned.

Answers in Section 1.5

1.4.2 Primary, Secondary, and Tertiary Prevention

The three examples of prevention that we have just discussed are (respectively) illustrations of ***primary***, ***secondary***, and ***tertiary*** prevention. These terms can be defined as shown in Table 1.4.1.

Table 1.4.1 Primary, secondary, and tertiary prevention.

Term	Definition	Example studied
Primary	Preventing an infection, injury, or other disease process from occurring	Limiting vehicle speeds reduces the likelihood of a young person or child being involved in a road accident; if an accident does occur, the risk of serious injury or death from this cause is reduced.
Secondary	Early detection of a disease process at a stage where the course of the disease can be stopped or reversed	Offering mammography to the population of women aged 50–64 years allows earlier detection of breast cancer and a better chance of successful treatment.
Tertiary	Management of a condition that is already established in such a way as to minimise consequences such as disability and other complications	The diabetic patient generally requires lifelong treatment, and the underlying condition cannot be cured. The complications and disability arising from foot problems, for example, can be avoided or ameliorated by an active approach to prevention.

Self-Assessment Exercise 1.4.2

For each of the following activities, state whether this is primary, secondary, or tertiary prevention, giving brief reasons for your answer:

1. Measles immunisation
2. Smear tests for cervical cancer every 5 years
3. A well-managed programme of terminal care for a patient with cancer
4. Use of bed nets impregnated with insecticide in malaria-endemic areas
5. Smoking-cessation programme in middle-aged men recovering from a heart attack

Answers in Section 1.5

1.5 Answers to Self-Assessment Exercises

Section 1.1

Exercise 1.1.1

The uses of epidemiological methods and thought: This list is not necessarily exhaustive, but it covers the most important applications:

- By studying populations rather than those already in the health-care system, one can gain a more-representative and complete picture of the distribution of disease. Studies of populations and those receiving care can identify the factors that determine who does, and does not, take up health care, and why.
- Describing the frequency of a disease, health problem, or risk factor; who is affected, where, and when. This may be used for ***planning***, as in epidemic control, service provision, etc.
- Understanding the ***natural history*** of health problems; that is, what happens if there is no treatment or other intervention.
- Understanding the ***causes*** of disease, thus laying the basis for prevention.

- Determining the ***effectiveness*** of health interventions, whether drugs, surgical operations, or health promotion through, for example, raising awareness or establishing public policy.
- Through an understanding of the determinants of the health of populations, epidemiology contributes to the development of ***prevention policy***.

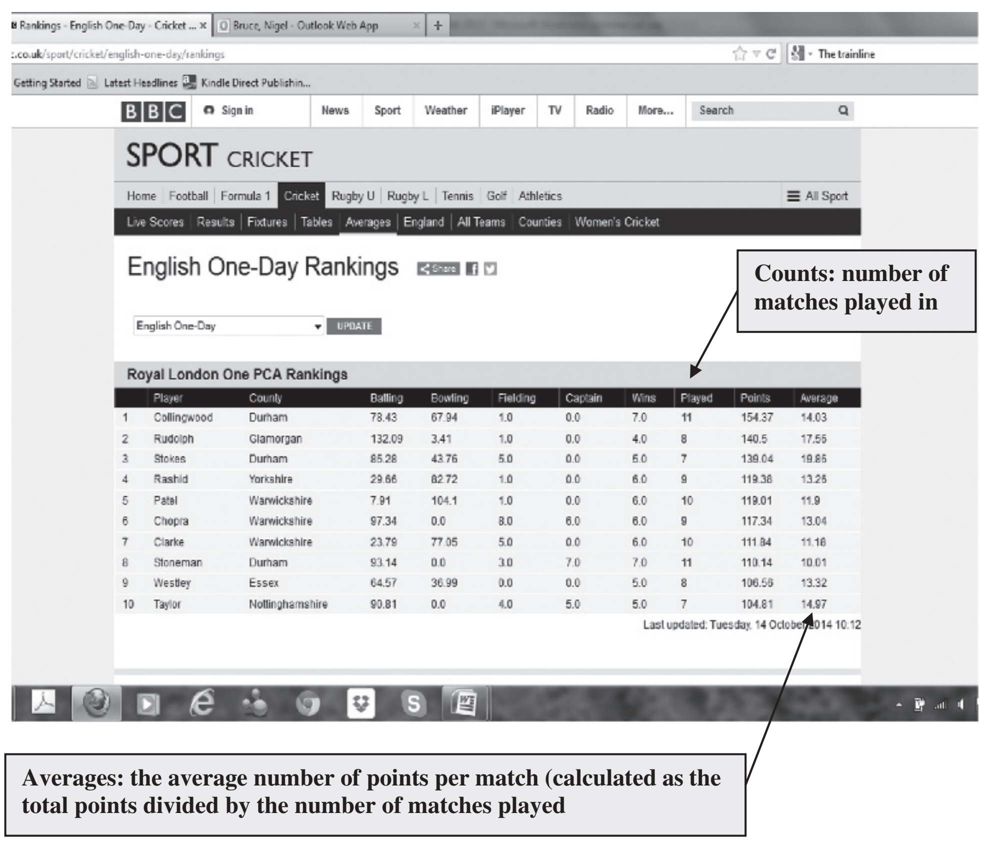
> Thus, while basic research may add to our biologic understanding of why an exposure causes or prevents disease, only epidemiology allows the quantification of the magnitude of the exposure–disease relationship in humans and offers the possibility of altering the risk through intervention. Indeed, epidemiologic research has often provided information that has formed the basis for public health decisions long before the basic mechanism of a particular disease was understood.
>
> *Hennekens and Buring, 1987, p. 13*

> Study of groups that are particularly healthy or vulnerable is often the beginning of the search for causes, and so of prevention.
>
> *Morris, 1957, p. 263*

Exercise 1.1.2

These are just a few examples of the use of statistics found on the Internet. Note the different ways that the information is presented.

1. Cricket scores with examples of counts and averages (BBC Sport website)

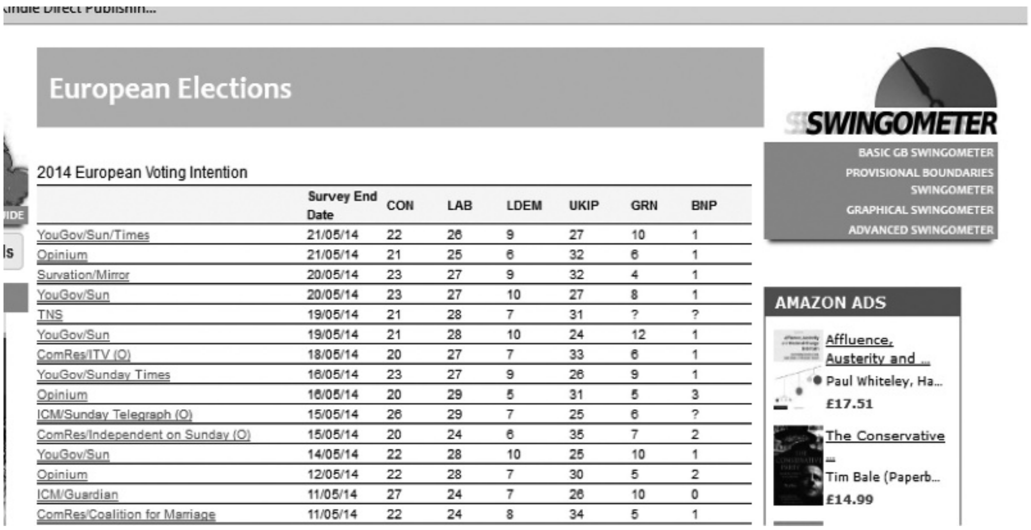


Royal London One PCA Rankings

	Player	County	Batting	Bowling	Fielding	Captain	Wins	Played	Points	Average
1	Collingwood	Durham	78.43	67.94	1.0	0.0	7.0	11	154.37	14.03
2	Rudolph	Glamorgan	132.09	3.41	1.0	0.0	4.0	8	140.5	17.56
3	Stokes	Durham	85.28	43.76	5.0	0.0	5.0	7	139.04	19.86
4	Rashid	Yorkshire	29.66	82.72	1.0	0.0	6.0	9	119.38	13.26
5	Patel	Warwickshire	7.91	104.1	1.0	0.0	6.0	10	119.01	11.9
6	Chopra	Warwickshire	97.34	0.0	8.0	6.0	6.0	9	117.34	13.04
7	Clarke	Warwickshire	23.79	77.05	5.0	0.0	6.0	10	111.84	11.16
8	Stoneman	Durham	93.14	0.0	3.0	7.0	7.0	11	110.14	10.01
9	Westley	Essex	64.57	36.99	0.0	0.0	5.0	8	106.56	13.32
10	Taylor	Nottinghamshire	90.81	0.0	4.0	5.0	5.0	7	104.81	14.97

Last updated: Tuesday, 14 October 2014 10:12

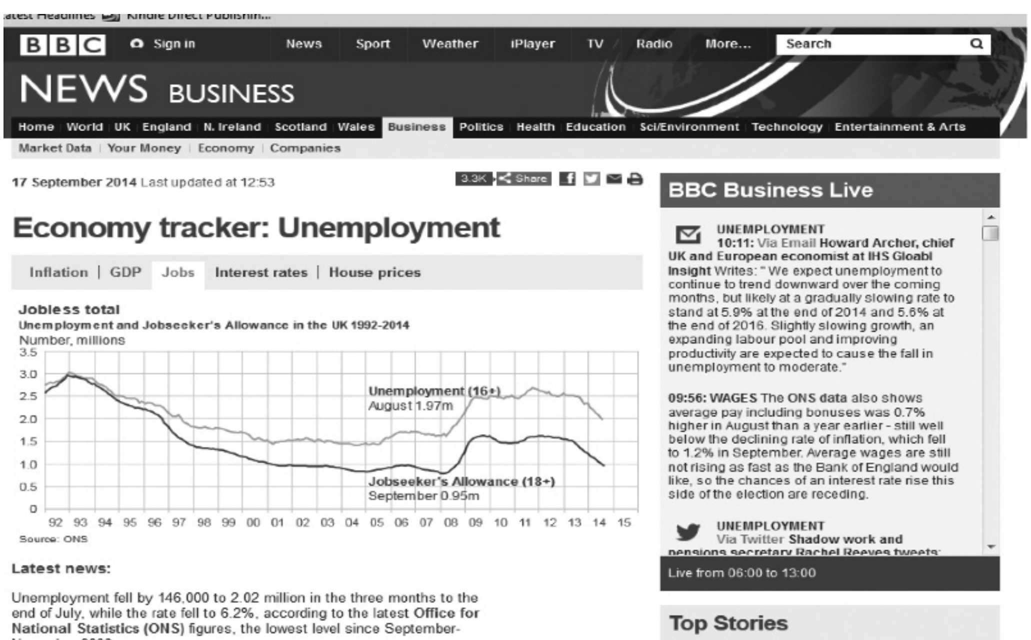
2. Voting intentions from opinion polls for the 2014 European elections, showing examples of percentages (*Source:* YouGov)



2014 European Voting Intention

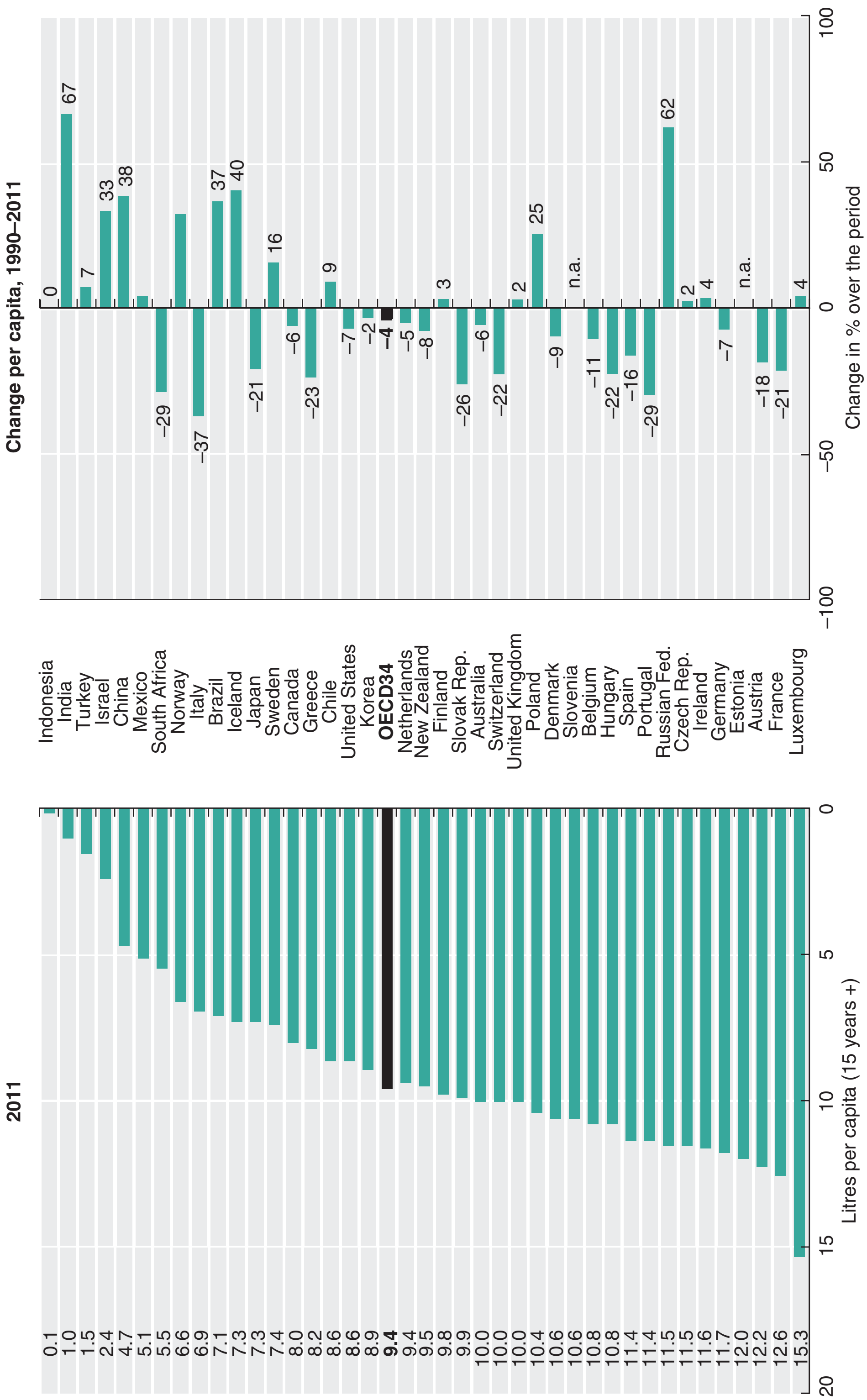
	Survey End Date	CON	LAB	LDEM	UKIP	GRN	BNP
YouGov/Sun/Times	21/05/14	22	26	9	27	10	1
Opinium	21/05/14	21	25	6	32	6	1
Survation/Mirror	20/05/14	23	27	9	32	4	1
YouGov/Sun	20/05/14	23	27	10	27	8	1
TNS	19/05/14	21	28	7	31	?	?
YouGov/Sun	19/05/14	21	28	10	24	12	1
ComRes/ITV (O)	18/05/14	20	27	7	33	6	1
YouGov/Sunday Times	16/05/14	23	27	9	26	9	1
Opinium	16/05/14	20	29	5	31	5	3
ICM/Sunday Telegraph (O)	15/05/14	26	29	7	25	6	?
ComRes/Independent on Sunday (O)	15/05/14	20	24	6	35	7	2
YouGov/Sun	14/05/14	22	28	10	25	10	1
Opinium	12/05/14	22	28	7	30	5	2
ICM/Guardian	11/05/14	27	24	7	26	10	0
ComRes/Coalition for Marriage	11/05/14	22	24	8	34	5	1

3. Unemployment data for the UK 1992–2014, showing percentage unemployment and those claiming the job-seekers' allowance, presented as line graphs



4. Alcohol consumption (litres per capita) among adults by country in 2011 and change over the period 1990–2011 (as a percentage), presented as bar charts (*Source:* Organisation for Economic Co-operation and Development)

JWST868-c01 JWST868-Bruce October 25, 2017 9:56 Printer Name: Trim: 254mm × 178mm



Section 1.2

Exercise 1.2.1

This is a complex problem, so don't worry if you found it challenging. Many different research questions could arise from this, depending on the type of research your experience and interests relate to. For example, at one end of the spectrum, a laboratory-based scientist might wish to carry out some physiologically based animal experiments involving chemicals from traffic pollution, food additives, and fungal spores (from damp housing). Laboratory-based experimentation with humans is out of the question (apart from specific volunteer studies), so we will concentrate on epidemiological investigation. The following two research questions represent different levels of investigation. The first aims to start the research process off by describing the situation (from which associations between asthma and other factors can be investigated), and the second takes a more analytical approach. Which one is appropriate would depend on exactly what has already been studied, resources available, and so on. You will note that both questions define the population as children aged 5–15 years, in order to help focus the study.

Research Question 1

How are asthma, levels of air pollution, damp housing, consumption of [specified] processed foods, and socioeconomic circumstances distributed among children aged 5–15 years living in [named city]?

Research Question 2

Among children aged 5–15 years living in [named city], are the presence and/or frequency of asthma associated with [specified] processed foods or damp housing, after taking account of levels of air pollution?

Section 1.3

Exercise 1.3.1

The reasons for this very large increase could be as follows:

- A ***chance*** (random) variation (although this is unlikely given the large numbers involved). The role of chance variation is a very important concept in epidemiology and statistical methods, and we will begin to examine this in Chapter 2.
- An ***artefact*** of the system, such as a change, or error, in the system for recording admissions. This is also unlikely to cause such a dramatic increase, but it needs to be considered. The quality of information and how it is defined and handled are very important issues in research, and we will begin to look at these in Chapter 2.
- A ***real*** increase, which could result from changes in referral procedures by GPs (although this is again unlikely to cause such a large increase in one year, especially given the more gradual increase seen in previous years). A more likely explanation is a sudden increase in the population that the accident and emergency department is serving. Natural increase in population is unlikely, unless there had been an event such as a rapid influx of refugees for example. Closure of another (smaller) accident and emergency department in the city is the most likely reason for the large increase in the population being served.

The key point here is that in order to make some judgment about this increase, we need to know the size of the population from which the admissions are coming. If this population has increased by 30 per cent, then, all other things being equal, we would expect the number of admissions to increase by 30 per cent. Changes in numbers of events (whether deaths, cases, admissions, etc.) cannot be interpreted usefully without information on changes in the population from which these events arose and the time period concerned.

Exercise 1.3.2

1. Calculation of point and period prevalence. ***Point prevalence***: In this diagram, 7 cases were 'active' at time A. With a population of 1,000, the point prevalence is 7 ÷ 1,000 = 0.007. This is a rather untidy way of expressing the prevalence, so we can state it as 0.7 per cent or 7 per 1,000. ***Period prevalence***: Here we have a longer period to consider, during which some cases resolve and new ones occur. We include the cases that got better, because they were cases at some time during period B. We must also include the new cases that start during period B. A total of 10 cases were active during period B, so the period prevalence is 10 ÷ 1,000 = 0.01. This can be presented as 1 per cent or 10 per 1,000 over the period concerned (e.g. 1 year).
2. The point and period prevalence rates differ because, for period prevalence, we included some cases that had not yet recovered and some new cases that appeared during period B.

Exercise 1.3.3

1. During period B, a total of seven new cases arose.
2. The incidence rate for boys aged 0–14 years is 78 ÷ 41,000 × 10,000 per year = 19.0 per 10,000 per year. The incidence rate for girls aged 0–14 years is 76 ÷ 37,100 × 10,000 per year = 20.5 per 10,000 per year. So although there were more cases among boys, the incidence rate was actually higher among girls. The reason for this is that the population of girls was smaller. This again emphasises why rates are so important.

Exercise 1.3.4

1. The total observation time for the 30 subjects is 261.9 person-years, during which time $\boldsymbol{n}$ = 20 subjects developed the disease. The incidence rate is 20/261.9 = 0.0764 per person-year or 76.4 per 1,000 (10^3) person-years.
2. In this case the total observation time is:

$$5 + 5 + 5 + 3.5 + 4.8 + 4.6 + \ldots 5 + 1.1 + 1.5 = 115.8 \text{ person-years}$$

The total number of cases is now 11 (as we only include cases occurring within the first 5 years), so the incidence rate is 11/115.8 = 0.095 per person-year or 95.0 per 1,000 person-years. This rate is somewhat higher than that for the longer follow-up.

Section 1.4

Exercise 1.4.1

Road Accident Prevention

Reducing vehicle speeds reduces the likelihood of a road accident, and it also reduces the chance of serious injury or death if a collision (with a person, vehicle, or other object) does occur. The prevention process here is principally through preventing the accident (injury or death) happening in the first place, but if it does happen, the severity of injury and the likelihood of death are reduced.

Breast Cancer Prevention

From the information we have, it is apparent that we do not yet know enough to prevent most cases of breast cancer from occurring in the first place. What we can do is detect the disease at a stage where, if treated promptly, it is possible to cure the disease in a substantial proportion of affected women. Thus, in contrast to the road accident example, the prevention begins after the disease process has started, but it is at a stage where it is still possible to cure the disease.

Prevention of Foot Problems in the Diabetic Patient

In this last example, the disease process is well established, and it cannot be removed or cured. That does not mean that the concept of prevention has to be abandoned, however. In this situation, preventative action (education, support, treatment, etc.) is being used to prevent damage to the skin of the feet, with the infection and ulceration that can follow. Prevention activities are carried out, even though the underlying disease remains present.

Exercise 1.4.2

Example	Prevention approach	Reasons and explanation
Measles immunisation	Primary	Immunisation raises immunity of the recipient and prevents infection.
Smear tests	Secondary	The smear test is a screening procedure designed to detect the disease process at an early stage. Note that the test on its own is not prevention, as it does not alter the course of the disease; it must be followed up (if positive) by biopsy and treatment as necessary.
Terminal care	Tertiary	The patient is dying (e.g. from cancer), and nothing can be done to alter that. That does not mean the preventative approach is abandoned. Good terminal care can prevent a lot of pain and emotional distress in both the patient and the patient's family and friends. This can have lasting benefits.
Bed nets	Primary	Impregnated bed nets prevent contact between the feeding mosquito and humans, especially at night, when most biting occurs. This prevents introduction of the malaria parasite into the body, thus preventing the occurrence of the disease.
Smoking cessation	Tertiary?	This one does not fit definitions easily. Smoking cessation in a healthy younger person could properly be regarded as primary prevention of heart disease. In our example, the men had already had a heart attack. This is not secondary prevention (detection at early stage, etc.). It is probably best seen as tertiary, especially bearing in mind that this smoking cessation should generally be part of a broader rehabilitation package helping the man to get mentally and physically better, reduce his risk of another heart attack, and get back to work or other activity.