CHAPTER

17 Scientific investigations

KEY KNOWLEDGE

This chapter is designed to enable students to:

- understand scientific research methodology and the typical steps in a scientific approach
- design and conduct an experimental or an observational investigation that follows scientific process
- recognise that investigations must adhere to relevant safety, health and bioethical guidelines
- distinguish between quantitative and qualitative data sets and to assess their validity, accuracy, precision and limitations
- become familiar with the various means of representing data
- recognise the nature of scientific evidence and its distinction from opinions and anecdotes
- understand that conclusions from a data evaluation may either support or disprove a hypothesis, model or theory
- recognise characteristics of effective communication of scientific ideas.

FIGURE 17.1 Scientific investigations may be carried out in a laboratory or in the field. Here we see students performing a field-based investigation under the guidance of a teacher. Note that one student is using an electronic notebook to keep a record of observations.

Knowledge from investigations

During your studies in biology, you will develop knowledge and an understanding of many concepts in new areas of study including cell structure and function, homeostatic mechanisms that enable survival, relationships in ecosystems, Mendelian genetics, and human reproduction and development. In addition to building knowledge and understanding in these specific areas of study, you will also develop an understanding of:

1. the scientific processes by which valid explanations for biological phenomena are formed and evaluated
2. the avenues of communicating biological knowledge to others.

In the following sections, we will explore the processes by which explanations in biology are developed.

Advancing knowledge in biology

How is our understanding of aspects of biology achieved? How do refinements of existing theory come about? Advances in biological knowledge typically begin with observations that raise questions for investigation, through avenues such as:

- **Carefully planned laboratory-based or field-based experiments designed to support or refute a particular hypothesis.** For example, the laboratory-based experiments of Oswald Avery during the 1940s identified the genetic material as being DNA (refer to figure 13.34 in the text) and refuted the proposal that the genetic material was protein.
- **Critical reinterpretation of previously accepted facts, producing a new framework.** For example, reptiles, such as snakes and lizards, are commonly believed to be thermo-conformers; that is, their body temperatures vary passively with the temperature of their external environments. However, studies have shown that several species of lizard are able to control their body temperatures regardless of changes in the temperature of the external environment. These lizards are thermo-regulators. Interestingly, they only regulate their core temperature when the cost of producing the necessary metabolic heat is not too high. If the cost is too high, these lizards simply let their body temperatures fall. Strictly, such reptiles are termed facultative thermo-regulators. These findings mean that the issue of temperature control in snakes and lizards is now known to be more complex than previously believed and is being reinterpreted.
- **Collection and analysis of new data.**
  - Example 1: Gregor Mendel's experimental crosses with sweet peas demonstrated that his factors (genes) were discrete particles that retained their identity across generations. This finding replaced the older concept that the factors involved in inheritance were non-particulate and lost their identity over generations through a blending process (refer to figure 13.31 in the text).
  - Example 2: The results of experiments carried out by T. H. Morgan led to the replacement of the view that Mendel's factors (genes) were free-floating particles that were always inherited independently. By analysing the offspring of crosses of specific fruit flies, Morgan and his team provided the first experimental evidence that genes were not free-floating particles but were physically located on chromosomes (refer to figure 14.26 in the text).
- **Identification and exploration of patterns or anomalies.** For example, the careful and independent observations of the German pediatrician Widukind Lenz (1919–1995) and the Australian gynecologist William McBride (1927–) identified the causal relationship between ingestion of the drug thalidomide by women during the first trimester of their pregnancies

ODD FACT

Thalidomide was withdrawn in late 1961. In 2008, thalidomide was approved for use, under rigorously controlled conditions, in the treatment of multiple myeloma and complications of leprosy.

and the occurrence of severe birth defects in their babies, including unusual limb malformations — a condition now termed thalidomide embryopathy (see figure 17.2 in this chapter and also refer to figure 12.7 in the text). Thalidomide came into clinical use, firstly in Germany in 1957, as an extremely safe tranquiliser, as a painkiller, and as an effective treatment for morning sickness in pregnant women. Thalidomide was quickly adopted for use in other countries, including the United Kingdom and Australia. The safety of thalidomide was based on the fact that overdoses caused extended sleep but did not cause death, which could occur with overdoses of other tranquillisers.

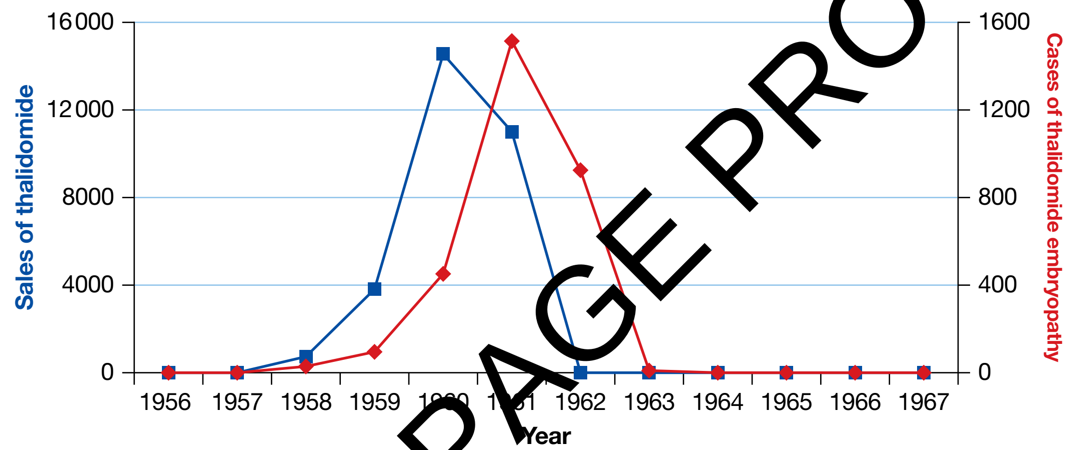


FIGURE 17.2 Line graphs showing data on the sales of thalidomide in Germany (blue dotted line) and the appearance of newborns with strange congenital malformations of the limb (red solid line), a condition that is now termed thalidomide embryopathy. Note that the pattern of appearance of malformed newborns follows the pattern of sales of thalidomide, but with a time lag of about nine months. The vertical axes denote sales of thalidomide (at left) and number of cases of malformed newborns (at right). (Data based on Lenz (19[illegible]4))

FIGURE 17.3 Photo showing Sir Alexander Fleming in his laboratory. Note the stacks of Petri dishes in the corner of his workbench.

You should note that not all advances in biological knowledge are initiated from the results of carefully planned experiments. Some major advances in biology are initiated by unplanned chance events, such as a discovery in 1928 by Sir Alexander Fleming (1881–1955). Fleming was studying *Staphylococcus* bacteria that he was growing on nutrient agar in petri dishes. Some of his plates became contaminated by the growth of airborne mould spores. Instead of discarding these plates as contaminated and useless, Fleming recognised that the clear areas around the growing mould were a signal that the mould was producing a substance that inhibited bacterial growth — indeed it was, and this substance turned out to be the antibiotic penicillin.

Investigations in biology

Carefully planned investigations are one process by which biological knowledge is advanced. These investigations may be either:

- experimental studies that are carried out
  - in a laboratory
  - in the field
- observational studies.

Both experimental and observational studies generate data that can be analysed, from which conclusions can be drawn. What is the difference between experimental and observational studies?

In an experiment, an investigator deliberately exposes a group of subjects or objects of interest to a defined treatment, with the aim of observing the group's responses to that treatment. For example, biologists might investigate the effect of various dosages of a new drug on a species of pathogenic bacteria. To do this, cultures of the particular bacteria would be grown in the presence of different dosages of the drug (the treatment), and the effect on bacterial growth (the response) at each level of the drug would be observed and recorded.

FIGURE 17.4 Biologists observing and recording the growth of a bacterial culture in the presence of particular levels of a new antibiotic drug

Figure 17.5 shows another experimental study in which a biologist is investigating the growth of a plant species under different environmental conditions.

In an observational study, an investigator collects data about the object of study but does not change the existing conditions. For example, a scientist might gather samples of water on a regular basis over several months from a river near an industrial site. These samples would be taken to the laboratory for analysis, to determine the concentrations of heavy metals in the river water over time (see figure 17.6). Refer to figure 3.30 in the text to see an example of scientists carrying out an observational study of the microbial life forms in the Movile Cave.

Other examples of observational studies are investigations carried out by biologists who gather data on the diversity and/or the abundance of species in a particular habitat. These studies often make use of quadrats at random locations along a transect line. Transects and quadrats are two ecological tools that enable biologists to quantify the relative abundance of one species or the biodiversity in an area such as a rainforest or a rocky shoreline or a sand dune.

FIGURE 17.5 A biologist examining the effect of a particular experimental treatment on the growth of seedlings.

FIGURE 17.6 An investigator collecting a sample of river water as part of an observational study of water quality at that site in the river. Why is this an observational study? Analysis of water samples collected on a regular basis over a period of time provides data on water quality at this point: for example, dissolved oxygen levels, concentrations of heavy metals or density of aquatic larvae.

Studies of changes in species abundance and biodiversity are important in tracking changes over time. Figure 17.7 shows an ecologist identifying and recording each species within a quadrat as part of an observational study of the recovery of a rainforest in northern Queensland after damage by a cyclone. Do you expect that this study would be repeated over intervals of time? In this case, the ecologist is recording her data in a notebook. As well as note-taking, other means of recording data include drawings, photographs and/or photographs. Refer to figure 8.63 in the text to see an example of students recording the abundance of an oyster species by sampling quadrats at several locations along a rocky shoreline.

FIGURE 17.7 An ecologist identifying and recording young plant species located within one quadrat as part of an observational study of the recovery of an area of rainforest in far north Queensland. Note the transect line on which the quadrat is located. The ecologist will use the quadrat to sample at several random locations along this transect line. Why sample at multiple sites rather than just one?

In the following section, you will be introduced to an observational study that ultimately led to major reforms in public health, including water supply and sewage.

The 1854 cholera epidemic

Cholera is an acute diarrhoeal disease caused by the *Vibrio cholerae* bacteria, which spread in water or food contaminated with the faeces of affected individuals. If left untreated, cholera can rapidly cause death, even within hours of infection. Severely affected individuals require rapid administration of intravenous fluids, up to six litres per day, and antibiotic treatment. Refer to chapter [illegible] in the text, under the section titled 'When transport goes wrong…' for details of how these bacteria cause the production of large volumes of watery diarrhoea. Many cholera epidemics and pandemics have occurred and some examples are shown in table 17.1 below. The World Health Organization estimates that, each year, three to five million people suffer from cholera, resulting in 100 000 to 120 000 deaths in developing countries.

TABLE 17.1 Some cholera outbreaks since the start of the nineteenth century. Note that not all countries affected in the various pandemics are shown. Estimates of the number of cases are available only for more recent outbreaks.

Date	Country/City	No. of cases	No. of deaths
Oct 2010 – Aug 2015	Haiti	700 000	9 000
Jul–Aug 1994	Congo	48 000	12 000
Jan 1991–Sep 1994	Peru	<1 000 000	10 000
1881–1896 fifth pandemic	Hamburg (1892)		8 600
1852–1860 third pandemic	Russia London (1853–54) Spain (1854–55)		>1 000 000 10 739 236 000
1829–1851 second pandemic	London (1849) London (1832–) Paris (1832–)		14 137 6 536 20 000

eBook*plus*

Weblink
A cluster is an aggregation of cases over a particular period in a given area without regard to whether the number of cases is more than expected.

Cholera had been raging through London during 1853 and the first half of 1854. At the end of August 1854, a localised outbreak of cholera began in the Soho district of London. Within three days, 127 deaths from cholera were recorded in Soho, and, in less than ten days, about 500 people were dead from cholera. A total of 616 deaths occurred during the two weeks of this outbreak. John Snow (1813–1858) was a doctor who lived close to the Soho district. Snow became aware that the cholera deaths were not uniformly distributed within Soho but were clustered in one area. Snow prepared a map showing the pattern of occurrence of cholera deaths within the Soho district. Figure 17.8 below is one representation of this map. Examine this map and note that the cholera deaths form a cluster that is centred around one particular pump, the Broad Street pump, marked with a yellow flag.

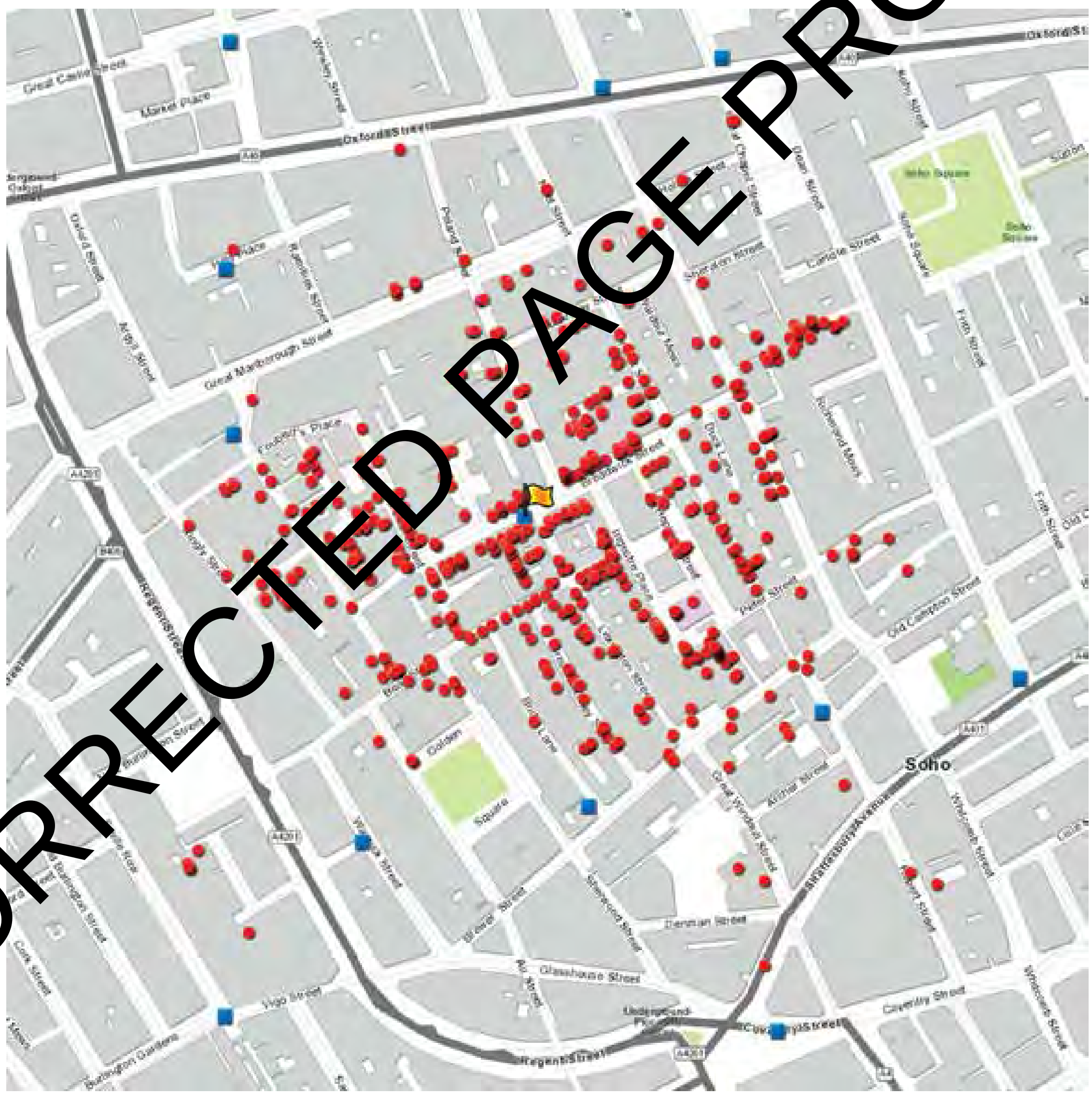


FIGURE 17.8 Representation of John Snow's map showing the deaths from cholera (red dots) in the Soho district of London in September, 1854. Also shown are the 13 street pumps (blue dots), with the Broad Street pump also marked by a yellow flag. Note the clustering of cholera deaths around this particular pump. Local residents collected water from these pumps for drinking and other purposes. (Image courtesy of Dr Don Boyes, Department of Geography and Planning, University of Toronto.)

In 1854, the prevailing medical and public view was that diseases were caused by bad or foul-smelling air filled with particles from decomposed matter. This view was the **miasma theory of disease** that was eventually replaced in the late nineteenth century by the **germ theory of disease**.

The Soho district was indeed full of foul-smelling air that came from rotting garbage and animal wastes in the streets, and from human faeces that, because of the lack of a sewage system, were typically emptied into cesspits or cess-pools either under houses or in streets. Houses in Soho did not have running water. The water supply in the Soho district consisted of shallow public wells from which water could be pumped. In all, 13 pumps were located around the Soho district. People took containers to these pumps to get the water they needed for drinking, cooking and washing. Figure 17.9 shows a cartoon of a typical nineteenth century public water pump.

FIGURE 17.9 Cartoon showing children playing at a street water pump typical of those in the Soho district of London in 1854. Note the handle whose action pumps water from a shallow underground well below the pump.

The pattern of deaths clustered in one region of Soho raised the initial research question for Snow: *What was the cause of the clustering of cholera deaths in Soho?* In addition, ignoring the miasma theory of disease, Snow also wondered: *Could water from the Broad Street pump somehow be involved in the cholera deaths?*

To investigate his research question, Snow asked more questions, made observations, and gathered data concerning instances of individuals who contracted cholera and who were known to have drunk water from the Broad Street pump, such as the following:

- nine customers of a small eating place where glasses of water from the Broad Street pump were served with other refreshments
- sixteen workers from a factory in Broad Street where tubs of water from the Broad Street pump were kept on hand for the workers to drink
- residents of houses close to the Broad Street pump who used that pump for their water supply.

In addition, some individuals who lived a distance away from the Broad Street pump also died from cholera, but Snow found a link to the water from that pump to the following cases:

- Members of some Soho families who lived much closer to pumps other than the Broad Street pump died from cholera. Did this observation mean that water from this pump was not linked to the cause of cholera? In some cases, these deaths were of children who walked from their homes to attend a school near Broad Street and, as they passed the pump, the children sometimes stopped to have a drink of water. In other cases, Snow identified that cholera deaths occurred in families who preferred the Broad Street water and chose to use that more distant pump for their water supply.

- One interesting case was that of an elderly woman who lived in West Hamstead, a cholera-free region at a distance from Soho. She died from cholera on 1 September. At first, this death appeared to have no link to the Broad Street pump. However, it was revealed by the woman's son that his mother had once lived in Soho and liked the taste of water from the Broad Street pump, so much so that she arranged for the daily delivery of a bottle of water from that pump. The woman received her last delivery of bottled water on 31 August, and she and her niece, who was visiting on that day, both died from cholera the following day.

Just as important was the data gathered by Snow about people who lived or worked close to the Broad Street pump who were ***not*** infected by cholera. This was necessary to exclude other possible sources of cholera infection. These examples included the following:

- A workhouse located very close to the Broad Street pump had more than 500 paupers in residence there. Very few deaths from cholera occurred among the inmates of this workhouse. This low figure was in marked contrast to the high death rates from cholera of people living in adjacent streets. Snow discovered that the workhouse had its own well from which water was pumped for use and, in addition, bought water from the Grand Junction Water Works, which sourced water from an uncontaminated region of the Thames River.
- Likewise, a brewery near Broad Street employed a large number of workers, but none died from cholera. Snow observed that the brewery workers had a daily allocation of beer and so drank beer rather than water. In any event, the brewery had its own well. (The fermentation processes involved in beer production kills cholera bacteria.)

Snow noted that there were no deaths from cholera among those people who did not drink water from the Broad Street pump. At the same time, he observed that everyone in Soho who died from cholera drank water from the Broad Street pump.

Snow rejected bad air (miasma) as the cause of the cholera deaths in Soho. Instead, Snow proposed that water from the Broad Street pump was the probable source of the cholera infection that raged in Soho. Based on this, Snow then refined his original research question into a working hypothesis that could be tested. His hypothesis was as follows: ***'IF water from the Broad Street pump is the cause of cholera deaths in Soho, THEN cholera deaths will decline and stop when this supply of water is stopped.'*** Note that Snow's hypothesis includes an explanation for the cholera deaths (water from the Broad Street pump) and a prediction by which the hypothesis can be tested (see if cholera deaths stop when this water supply is stopped).

Although local officials were sceptical and supported the miasma theory of disease, Snow persuaded them to remove the handle from the Broad Street pump, making it non-functional and stopping the supply of water from its underground well. The pump handle was removed on 8 September 1854 and the cholera deaths declined and stopped. However, the effect of removing the pump handle on the Soho cholera outbreak is confounded by the fact that, as may be seen in figure 17.10, by the time the handle was removed, the cholera outbreak was already declining. In any event, the relationship between the cholera deaths in Soho and drinking water from the Broad Street pump as hypothesised by John Snow was supported, and the removal of the pump handle certainly prevented any later outbreaks of cholera.

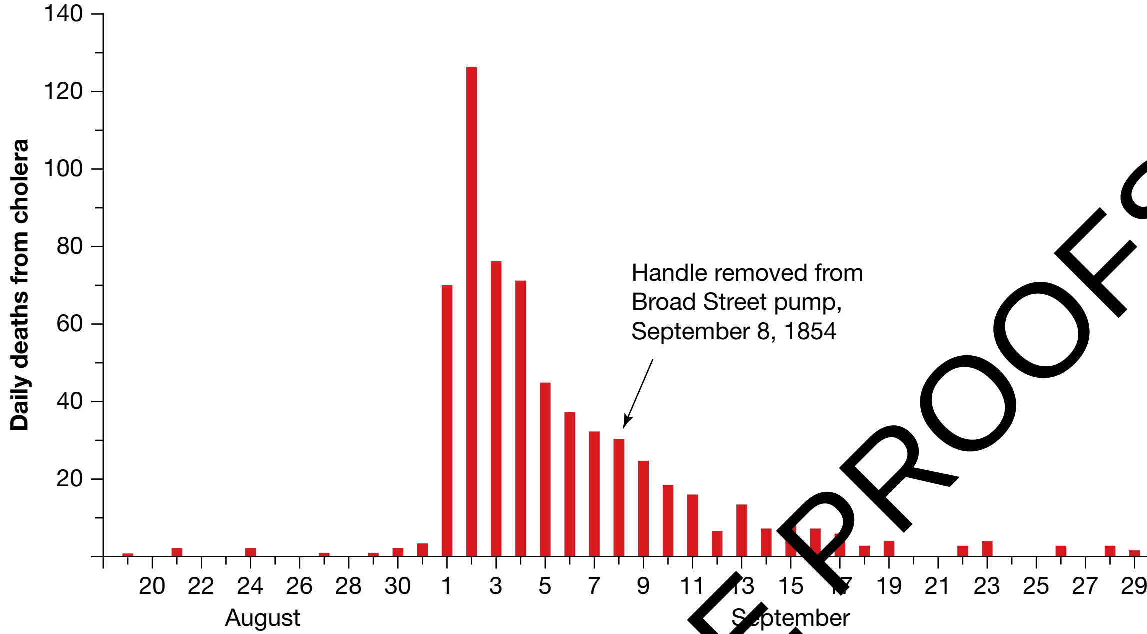


FIGURE 17.10 Bar graph showing the daily deaths from cholera in Soho over late August and early September 1854. Note that, by the time the pump handle was removed from the Broad Street pump, the local cholera epidemic was in decline.

John Snow's studies implicated the water from the Broad Street pump as the cause of cholera. He examined water samples from various pumps in Soho but was unable to demonstrate the presence of a causative agent of cholera in water from the Broad Street pump. This was not surprising. It was not until almost 30 years later, in 1883, that Robert Koch isolated pure cultures of *Vibrio cholerae* bacteria and demonstrated that these bacteria were the causative agent of cholera. Koch also showed that the mode of transmission of cholera was by ingestion of water or food contaminated with faeces from a cholera-infected person — a faecal-oral route — as happened in Soho when people drank water contaminated with faecal matter that leaked from a cesspool to the well supplying the Broad Street pump.

What was the origin of the Soho cholera outbreak? It is apparent that until August 1854, the water from the Broad Street pump was 'safe to drink' but, very suddenly, in late August, the situation changed and this water carried the causative agent of cholera. How did that water became contaminated? Evidence later emerged that, in late August, a five-month-old baby, Frances Lewis, developed severe diarrhoea, a symptom of cholera. The Lewis family lived at 40 Broad Street. Sarah, the baby's mother, related how she soaked the baby's diarrhoea-stained napkins in buckets of water and then poured the faecal-contaminated water into an opening into a leaky cesspool at the front of her house. However, this cesspool was just a metre from the well that fed the Broad Street pump. Baby Frances died from cholera on 2 September. Leakage of faecal material from the cesspool to the well under the Broad Street pump was apparently the trigger that set off the Soho cholera outbreak at the beginning of September, 1854.

eBook*plus*

Weblink
A cesspit or cesspool is a tank with no outlet used to store human faecal wastes.

Communicating research findings

The final step in Snow's investigation was the communication of the results of his research. Snow wrote many scientific papers dealing with cholera, including the following:

- 1849: On the pathology and mode of communication of cholera Part 1. *London Medical Gazette* 44, pp. 745–52.

- 1849: On the pathology and mode of communication of cholera Part 2. *London Medical Gazette* 44, pp. 923–29.
- 1851: On the mode of propagation of cholera Part 1. *Medical Times* 24, pp. 559–62.
- 1851: On the mode of propagation of cholera Part 2. *Medical Times* 24, pp. 610–12.
- 1853: On the prevention of cholera. *Medical Times and Gazette* 7, pp. 367–69.
- 1854: The principles on which the treatment of cholera should be based. *Medical Times and Gazette* 8, pp. 180–82.

Communicating research findings acts to share information with other groups, possibly leading to further research or, in some cases, resulting in new practices and new treatments. In the case of John Snow, his actions in Soho in identifying the Broad Street pump as the source of contaminated water that caused cholera was not the end of the story. Importantly, his research on cholera made a major contribution to public health by demonstrating the need for clean water supplies and sewage systems in cities and towns, and his findings contributed to the reforms that improved public health in the nineteenth century.

John Snow carried out an important observational study of cholera deaths in Soho using a scientific approach. As part of your biological studies, you are required to design and undertake an investigation using a laboratory-based study, a field-based experimental study and/or an observational study. Your investigation should relate to an aspect of the survival of an organism or a species, noting that survival typically involves the regulation of factors within an individual organism or species. You will also be required to communicate your findings.

In the following section, we will look at the process of designing and undertaking a scientific investigation and communicating the findings of that investigation.

ODD FACT

The *John Snow* pub on the corner of Broadwick and Lexington streets in London was named in honour of Snow. Broadwick Street is the renamed Broad Street of the nineteenth century.

KEY IDEAS

- Progress in science occurs when existing theories in science are modified or even rejected and replaced.
- Carefully planned investigations are one process by which biological knowledge is advanced.
- Scientific investigations may be experimental or observational.
- Experimental studies may be carried out in the laboratory or in the field.
- In experimental studies, investigators expose the subjects or objects of interest to a defined treatment and observe their responses to those treatments.
- In observational studies, investigators collect data about an object of study without changing the existing conditions.

QUICK CHECK

1. Put the following actions of John Snow in the correct order over time:
   - **a** formulating a testable hypothesis
   - **b** communicating results
   - **c** asking questions
   - **d** gathering data
   - **e** analysing results
2. State whether the following statements are true or false:
   - **a** John Snow's hypothesis was tested by carrying out an experiment.
   - **b** Cholera was once thought to be caused by bad air or miasma.
   - **c** Snow identified the bacterial species that is the causative agent of cholera.
   - **d** Advances in science may involve the rejection of an accepted theory.
   - **e** An observational study is completed by collecting data.

Process of scientific investigation

Figure 17.11 summarises a scientific approach to the investigation of a research question using laboratory-based or field-based experimental studies or using observational studies.

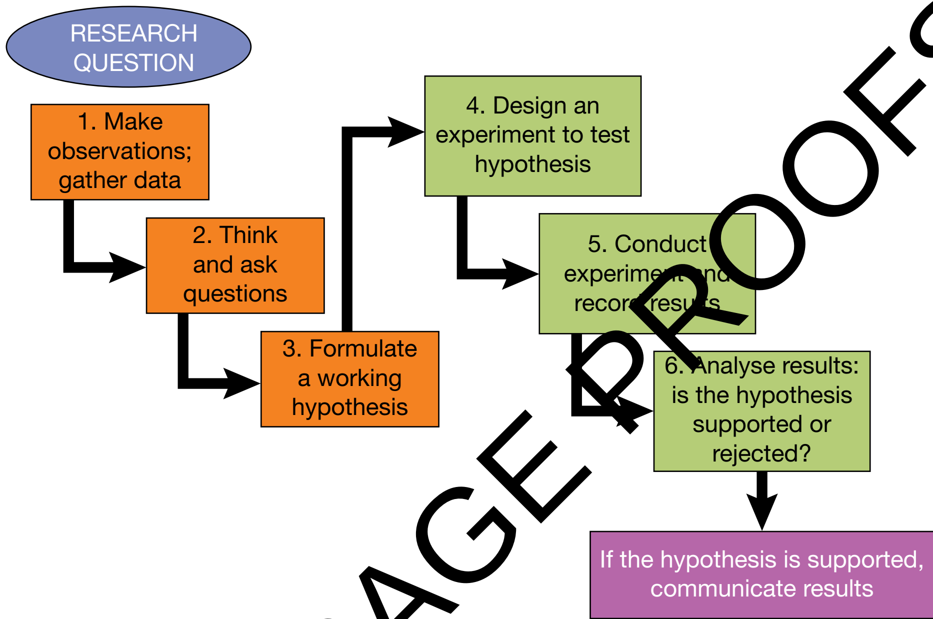
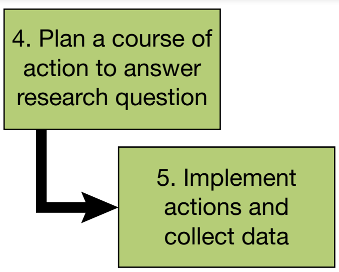


FIGURE 17.11 Diagram outlining typical steps in a scientific approach to a research question. Note that an investigation does not start with an experiment. It starts by gaining background information about the research question and by making observations and gathering data.



In the case of an observational study, steps 4 and 5 in figure 17.11 above would be amended as shown in the figure in the minor column.

Let's look at these steps in turn.

Choosing a research question

Every study, both experimental and observational, begins with a research question. In particular, a biologist might be thinking about questions such as:

- 'Is the water quality of a river near an industrial site the same as that upstream of the site?'
- 'Can a particular plant species survive under conditions of increasing salinity?'
- 'What relationship exists between light intensity and photosynthetic rate in plants?'
- 'Does humidity affect transpiration rates in a plant?'
- 'What animal behaviours contribute to the animal's survival in hot environments?'

For John Snow, his initial research question was: 'What was the cause of the clustering of cholera deaths in Soho?'

You should decide on a relevant question that you are interested in investigating through either an experimental or an observational study.

Once you have decided on your research question, you will use a scientific approach to investigate your question, as shown in figure 17.11 above and as discussed below. As you proceed through the early steps of the scientific process, your research question will be modified and will finally develop into a working hypothesis. Testing this hypothesis will be the focus of your laboratory or field experiment and/or your observational study.

Step 1. Observations produce data

Observation is a fundamental skill used by scientists to gain data — in other words, observations are a source of raw data. Observations provide the starting point of an investigation but, on their own, they do not provide explanations or answers.

FIGURE 17.12 Scientist viewing an image using a transmission electron microscope (TEM). Because its resolving power is orders of magnitude greater than that of light microscopes, a TEM can reveal remarkable details of cellular infrastructure.

Making observations and gathering data

Observations can be made directly through your senses, by using your eyes to record the green and yellow colours of the variegated leaves of a flowering plant, your hands to identify the rough texture of its stem, your olfactory sense to note the smell of its flowers, and your tongue to detect the sweetness of the fruit if it is edible and safe.

As well as the unaided senses, scientific observations are more often made using instruments that extend the power of the unaided senses, or permit accurate measurements, or enable collection of data that may otherwise be undetectable. Examples include thermometers to measure temperature, micrometers to record thicknesses, scales to measure mass, microscopes to visualise cells and cell organelles (see figure 17.12), DNA sequencers to identify the base sequences of genes, or magnetic resonance imaging (MRI) to make a medical diagnosis.

Not all observations are made using instruments in a laboratory setting. Familiar tools that can be used to gather data include digital audio recorders, video recorders and digital cameras. Tools are an important means of gathering data from people and these include questionnaires (oral or printed), surveys and interviews. Quadrats and transects are tools used in the field to gather data. Can you think of other tools for making observations and gathering data?

At the beginning of your investigation, you will gather data that provides background information relevant to your research question. You may, for example, gather some data using web searches. You may also gather other background information from books, journal articles and popular science magazines. You may talk to acquaintances or friends who you believe may have knowledge or suggestions about your research question. It is essential that you keep a logbook in which you record the activities you complete in your information-gathering sessions. For example, when you do a web search, note in your logbook the URL (Universal Resource Locator) of any websites from which you access data. See the box on logbooks at the end of this section.

Later in your investigation, your observations and data gathering will focus on recording the results of the experiments that you perform or the observations that you make in order to test your research hypothesis.

Sources of data

Data is a set of facts that are collected, observed, or generated. Data that you gather may come from primary sources or from secondary sources. **Typically, data that is gathered is raw data that must be analysed and interpreted to produce useful information.**

Primary sources of data provide direct or firsthand evidence about some phenomenon, as, for example, a research investigation, an event, an object, or community attitudes. Major sources of primary data are results of investigations reported in scholarly journals by the researchers involved, as,

eBook*plus*

Weblink

In this chapter, 'data' when used as a mass noun is singular, while 'data' used as a count noun is plural.

for example, research articles in the journal *Nature*. Other sources of primary data include video recordings of events, eyewitness accounts of events, historical documents, personal diaries, legal documents, completed questionnaires and surveys, and interview transcripts. Your completed logbook will be a primary source of data about your investigation of your research question.

Secondary sources of data are comments on or summaries and interpretations of primary data. Sources of secondary data include review articles in newspapers and popular science magazines, such as *New Scientist* or *Scientific American,* written by one person who summarises and comments on the research of others. A book about Charles Darwin is a secondary source, but a book containing transcripts of Darwin's letters is a primary source. (Why?)

Two kinds of data

Data that you gather may be quantitative or qualitative.

Some data that you gather will be recorded in numbers. For example, biologists might record a glomerular filtration rate of 132 ml per minute, a child's red blood cell count of 4.8 million per microlitre, a feverish body temperature of 41 °C, a transpiration rate of 0.25 mL per minute, a duration of the S phase of human cell cycle of about 8 hours. This type of data is termed quantitative. **Quantitative data can be both observed and measured and expressed numerically.** Can you think of other examples of quantitative data? Quantitative data can also be manipulated mathematically to produce useful statistics, such as means (averages) and standard deviations.

Other data that you collect will be expressed in word descriptions. For example, in an experiment, a student observed that treating the surface of a cut potato with iodine solution produced a purple colour. 'Purple' is an example of qualitative data. **Qualitative data is descriptive, not numerical, and can be easily observed and described, but not measured.** Other examples include descriptions of texture (rough, smooth, bristly, coarse, clammy, hairy, and so on) or smell (acrid, pungent, aromatic, sweet, musty, smoky, fragrant, and so on).

The results of laboratory experiments are most commonly quantitative data. In contrast, data gathered through the use of surveys, questionnaires and interviews are typically qualitative. What kind of data would you expect to gather if you used a survey-based investigation into community attitudes towards an issue, such as genetic testing, genetic screening, or the use of embryos as a source of stem cells?

Table 17.2 below shows examples of how some values of attributes can be expressed both qualitatively and quantitatively.

TABLE 17.2 Some attributes can be expressed either as qualitative or quantitative data. Which form provides more detailed information?

Attribute	Qualitative	Quantitative
height (person)	tall	183 cm
height (mountain)	very high	8848 m
colour	green	520 nm
sound	loud	85 dB
area (postage stamp)	small	8.75 cm^2
speed	fast	120 kms^{-1}
temperature	hot	100 °C
age (rock)	ancient	440 million years

Recording data

Scientists gather raw data or plain facts from their observations. Data must be recorded *at the time* in a suitable form, as, for example, text entries, sketches, tables, and/or diagrams in logbooks, or in field notebooks. These records may be supplemented by audio and video recordings.

FIGURE 17.13 Record keeping in a logbook is an important part of the scientific method.

KEEPING A LOGBOOK

A critical task for investigators is to keep a logbook (or field notebook). Your logbook may be a bound exercise book (not loose-leaf) or a digital document on your computer or iPad/tablet. It will contain all the details of your progress through the steps of your scientific investigation, starting from the process of deciding on your research question and through each step of your investigation, which will follow a scientific approach. Your logbook is an important and valuable document. Note that even famous biologists kept notebooks: for example, Charles Darwin kept copious records of his observations in notebooks. During his voyage on the *Beagle* in 1831–36, Darwin filled 19 field notebooks with observations and thoughts that he made during shore excursions. (To read about Darwin's voyage on the *Beagle*, see *Nature of Biology 2 Fifth Edition*, chapter 10, in the box titled *The travels of Darwin and Wallace*.)

Make sure your name is prominently displayed on the front page of your logbook along with the title of your research topic. Your research question will be a topic of interest to you related to an aspect of cellular processes or to an aspect of biological change and continuity over time.

Keep a record of every session of your investigation. Your record might include (1) the date of the session, and (2) a heading that identifies the step in your investigation.

Possible headings you might use could include:

- **Deciding on my research topic.** The starting point for your investigation will be to decide on your research topic. Briefly outline why you decided on a particular research question.
- **Background data.** Through observations, you will gather background data (facts) on your research topic that can be recorded in your logbook as written notes, sketches, diagrams and tables; you might supplement these records with other media, such as photos, video recordings or audio recordings.

  Be sure to record not only the data items themselves, but also their sources, so that, if necessary, you can easily locate and revisit them. Refer to the box titled *Referencing data* later in this chapter for further detail on data sources.

(continued)

- **Questions arising.** In your logbook, record questions that arise and any thoughts that you have about your investigation. For each question, write your initial thoughts about possible answers. It does not matter if you later find that an answer is not correct. Simply record your amended answers — that's all part of a scientific approach. Identify any problems that arise during your investigation.
- **Developing my hypothesis.** Remember that a hypothesis is a tentative explanation and not simply a prediction. Refer to the section in this chapter titled *Writing a hypothesis statement* for more detail.
- **Planning my experiment.** Outline your plan using the headings: Aim, Materials, Methods, Results, Analysis of Results, Conclusions.
- **Carrying out my experiment.** Observe and record your experimental results in an appropriate form, such as tables of values in the case of quantitative data, and word descriptions in the case of qualitative data. You can later convert tabular data into other presentations, such as bar graphs and line graphs that you will incorporate into a poster presentation.
- **Analysis and evaluation of my results.** Record your interpretation of the experimental results and decide whether they support or disprove your hypothesis. Record your conclusions.

Step 2. Asking questions about data

The analysis and interpretation of raw data produces useful information. John Snow's analysis of the data that he gathered and mapped concerning cholera deaths in Soho led him to ask the question: *Did cholera deaths occur in Soho residents who did not drink water from the Broad Street pump?* This was a key question, because if no such deaths occurred, this would become evidence to support the supposition that the cholera deaths in Soho were linked to drinking water from the Broad Street pump. This question enabled Snow to refine his initial research question and formulate a working hypothesis.

As you review the background information that you have gathered from various sources, questions will arise in your mind about your research question. Record these questions in your logbook. Think about these questions and discuss them with your teacher and your classmates. This process will assist you to develop tentative conclusions and to then formulate a hypothesis that you will test through an experimental or an observational investigation.

FIGURE 17.14 Students discussing some of the data forming part of the background information to their research question.

Step 3. Formulating a hypothesis

Formulating a hypothesis is an important step in the scientific process. What is a hypothesis?

Some facts:

- A hypothesis is a tentative or provisional explanation for an observation — that is, it is an attempt to explain an observed phenomenon.
- A hypothesis is a statement, not a question.
- A hypothesis must be testable using the results of experiments or observations.
- A hypothesis must be falsifiable by a simple test or observation.

We can combine all these features into a single definition as follows:

A hypothesis is a tentative, testable and falsifiable explanation for an observed phenomenon.

If it is not an explanation, it is not a hypothesis, but . . . any explanation (correct or not) can be a hypothesis if it can be tested by observation or experimentation.

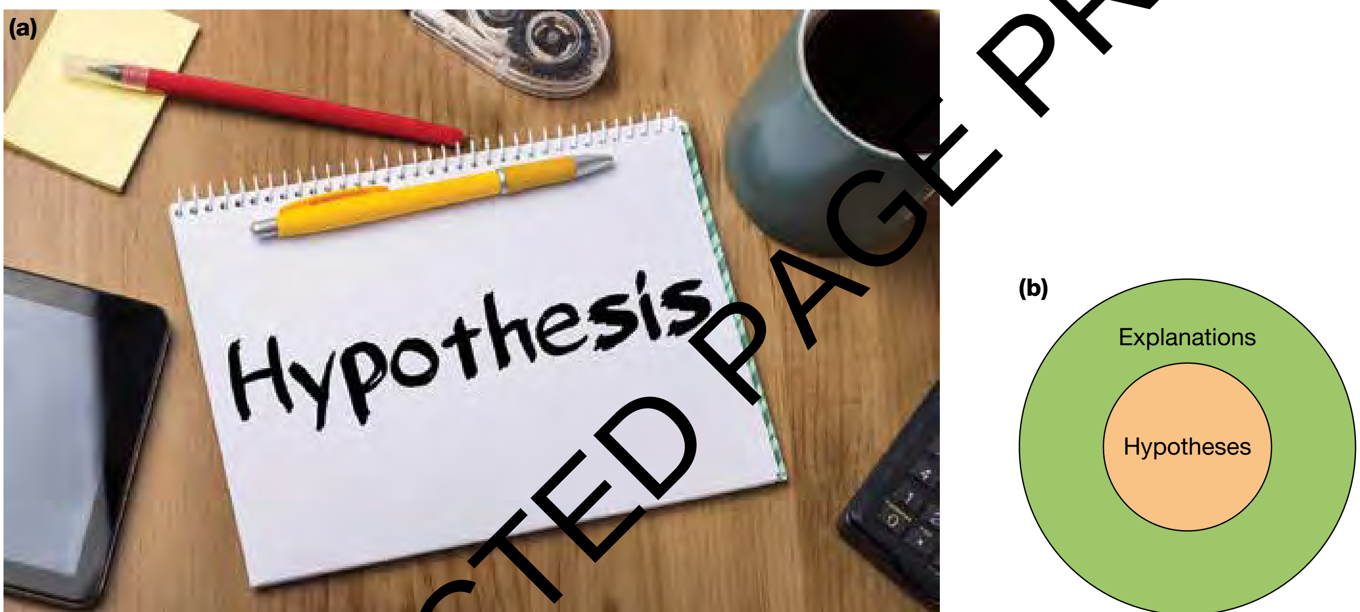


FIGURE 17.15 **(a)** Formulating a hypothesis is an essential step in a research investigation carried out using a scientific approach. **(b)** Venn diagram illustrating that all hypotheses are explanations of observed phenomena, but not all explanations are hypotheses. Only an explanation that can be tested and falsifiable is a hypothesis.

'**Testable**' means that a hypothesis can be easily tested by observations and/or experiments. Testing provides the evidence that allows a conclusion to be drawn about the hypothesis, namely that the hypothesis is either 'supported' or 'disproved (refuted).' **Scientific hypotheses are never stated as being 'proven correct'.**

A hypothesis can be any explanation for an observation *provided that it is testable*. However, not every suggested explanation can be tested. One example of an untestable explanation is the assertion that the disease known as childbed fever once common in new mothers was caused by 'unfavourable atmospheric-cosmic-terrestrial influences'. A testable hypothesis **cannot** be formulated from this. Try to complete the following statement: 'If unfavourable atmospheric-cosmic-terrestrial influences are the cause of childbed fever, then . . .'

'**Falsifiable**' means there must be a simple test, such as an observation, that could prove that a hypothesis is false. An example of a falsifiable statement is: 'All cats are black'. The testable prediction from this statement is 'If all cats are black and I see a cat, then it will be black'. This hypothesis can be readily falsified by a simple test — it simply requires seeing just one cat that is not black. However, in reality, if a hypothesis is not testable, it is also not falsifiable.

A hypothesis is a tentative, but testable, explanation for an observed phenomenon. For John Snow, the phenomenon that he observed was the cluster of cholera deaths centred on the Broad Street pump and he set about explaining this. In everyday life, you are often formulating simple hypotheses. You do this when you identify explanations for observations you make (see table 17.3).

TABLE 17.3 Examples of simple everyday hypotheses and how they can be tested.

Observation	Hypothesis	Test of hypothesis	Result	Conclusion
You turn on a torch and it does not work	Batteries are flat	Replace batteries and switch the torch on	Torch does not work	Hypothesis disproved
You are at the bus stop at the scheduled time, but no bus arrives	The bus is late	Wait and see if the bus arrives shortly	Bus arrives a few minutes later	Hypothesis supported
You turn on a torch and it does not work	The globe has blown	Replace the globe and switch the torch on	Torch works	Hypothesis supported
You become aware of the smell of burning	The toast is burning	Check toaster to see if the toast is burnt	Toast is black and smoking	Hypothesis supported
You hear the approach of blaring sirens	An ambulance is approaching	Look to identify the vehicle concerned	Police car with siren passes	Hypothesis disproved

Typically, a scientific hypothesis is formulated as an 'IF . . . THEN . . .' statement that starts with the tentative explanation and includes a prediction by which the hypothesis may be tested. In general, a good working hypothesis (regardless of whether or not it is later supported or is disproved) may have the form: 'IF hypothesis (explanation), THEN predicted outcome.' Table 17.4 shows some examples.

TABLE 17.4 Examples of some working hypotheses.

IF Hypothesis (explanation)	THEN Prediction
IF the rate of diffusion of molecules across a plasma membrane is directly related to their molecular size	THEN smaller molecules will cross the plasma membrane faster than larger molecules
IF skin cancers are caused by exposure to ultraviolet light	THEN people receiving higher UV exposures will develop more skin cancers

Hypotheses may also include an action and these have the general form: IF hypothesis (explanation) + action to be taken, THEN predicted outcome. For example:

- IF the torch batteries are flat and I replace them, THEN the torch will work again.
- IF the rate of photosynthesis is affected by the intensity of incident light and plants are illuminated with light of different intensities, THEN the rate of fixation of radioactive carbon by these plants will differ.

John Snow's hypothesis was: 'IF water from the Broad Street pump is the cause of cholera deaths in Soho, and water from this supply is turned off, THEN cholera deaths will decline and stop.'

Now consider the following statements and decide if each is an example of a working hypothesis:

Statement 1: 'Seedlings treated with organic fertiliser grow faster than similar untreated seedlings.'

No. This is simply a testable prediction. It does not include a tentative explanation.

Statement 2. 'Treating seedlings with organic fertiliser will make them grow faster than untreated seedlings.'

No. The statement is a method followed by a predicted outcome. It does not identify a tentative explanation for faster growth.

Statement 3. 'IF organic fertiliser provides essential growth factors for plants, THEN seedlings treated with fertiliser will grow faster than untreated seedlings.'

Yes. This identifies a tentative hypothesis (explanation) and a predicted outcome by which the hypothesis can be tested.

Variables in an experiment

A hypothesis usually predicts the relationship between two kinds of variables, an independent and a dependent variable. Let's look at the kinds of variables in an experiment to test a hypothesis.

In an experiment, a variable is any factor that the researcher can control or change or measure. Three kinds of variables are commonly recognised (see figure 17.16).

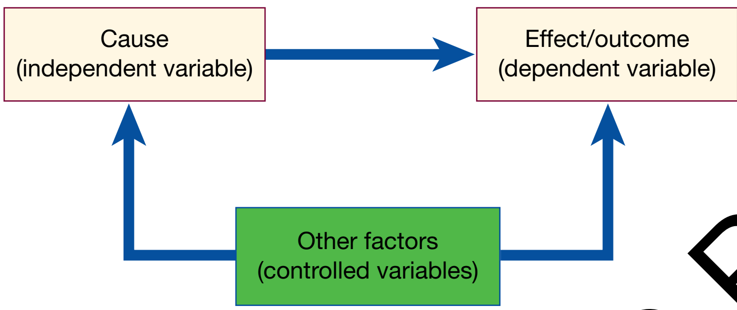


FIGURE 17.16 Diagram showing the relationships between variables in an experiment. Which interactions should you aim to prevent?

- An **independent variable** is a factor that is under the control of and is deliberately manipulated by the investigator. The investigator then observes and measures the effect of the manipulation of the independent variable on a dependent variable. For example, in an experiment investigating the relationship between the wavelength of light and the growth rate of seedlings, the independent variable that the investigator changes is the wavelength of the light. The dependent variable that is observed and measured in response is the rate of photosynthesis.

  In some studies, an investigator cannot manipulate an independent variable since it would be impractical or unethical to do so. For example, if an investigator wished to study the effect of an illegal drug on an aspect of human behaviour, it would clearly be unacceptable and unethical to ask one group of people to take such a drug. Research of this type might however be pursued using a suitable questionnaire administered to both non-drug users and drug users.
- A **dependent variable** is also termed an outcome variable. This is the factor that the investigator observes to see how it changes in response to changes in the independent variable. The changes in the dependent variable are measured and recorded. So, in the above investigation, the dependent variable is the rate of the growth of the seedlings.
- **Controlled variables** are all those other factors that the investigator must maintain at constant values through the course of an experiment. In the above experiment, the plant species used, the age of the seedlings, the temperature, availability of water, and angle of incident light are all examples of controlled variables. The controlled variables are ones that the investigator keeps the same throughout the course of the experiment. If these factors are not kept constant, they can confound the experimental results because they can cause changes in the dependent variable.

  Typically, an experiment involves one independent variable that is systematically changed, and one dependent variable whose responses to these changes are observed and measured. By changing only one variable (the independent variable) at a time, an investigator can see and measure the effect on the dependent variable.

In summary:
- the independent variable is what an investigator changes
- the dependent variable is what an investigator observes and measures
- controlled variables are what the investigator keeps constant.

Variables: categorical or numerical

The variables that are investigated in an experiment differ in their nature. They can be broadly divided into two classes — categorical variables and numerical variables. (These classes can be further subdivided, but this is not covered here.)

Categorical variables are qualitative and are expressed in words. A categorical variable can have a value that is one of a limited number of non-overlapping discrete values, and these values are expressed in words. Examples of categorical variables include the months of the year (twelve values: January to December), the days of the week (seven values: Sunday to Saturday), ownership of a smart phone (two values: Yes and No). Bar graphs are often used to display the frequencies of categorical variables in a study population.

Numerical variables are quantitative. Numerical variables can be precisely measured and they have values that are expressed in numbers. Examples of numerical variables include temperatures, masses, distances, rates of oxygen production in photosynthesis, wavelengths of light, salinity of water.

How would you classify these variables: breeds of dog? Air temperatures? Hair colours? Glomerular filtration rates of kidneys?

KEY IDEAS

- A scientific approach to an investigation is a multistep process.
- Observations produce data that may be quantitative or qualitative; the data must then be analysed to produce useful information.
- Sources of data may be primary or secondary.
- Both quantitative and qualitative data can be observed and described in words, but only quantitative data can be measured and given numerical values.
- Logbooks or field notebooks are used for the timely recording of progress through every step of an investigation.
- Hypotheses are tentative, testable and falsifiable explanations for observed phenomena.
- Hypothesis testing is a process whereby the result of an investigation provides evidence that either supports or disproves (refutes) a hypothesis.
- Variables in an investigation may be identified as independent, dependent or controlled.

QUICK CHECK

3 What is the difference between a primary and a secondary source of data?
4 What is the difference between a dependent and a controlled variable?
5 State whether the following are true or false:
   a Experiments are carried out only after observations relevant to the research question are made, questions asked about these observations and hypotheses formulated.
   b A research article in a scientific journal is an example of a primary source of data.
   c Green light and light with a wavelength of 520 nm are the same.
   d An independent variable is the variable that is systematically changed by the investigator.
   e The results of an investigation can prove that a hypothesis is true.

Step 4. Designing an experiment to test a hypothesis

Designing an experiment means constructing a detailed experimental plan to test a hypothesis in advance of doing the experiment.

A typical plan might include details under the following headings that you would record in your logbook:

Title of experiment:

Aim: Purpose of the experiment, which should include the hypothesis that you are testing.

Materials: This would include the following items that should be checked with your teacher:

- The laboratory equipment and the consumables needed for your experiment, any personal safety equipment required, and their availability checked.
- Any safety issues associated with the conduct of your experiment, such as:
  - potentially hazardous chemicals
  - equipment that can pose a potential safety hazard, as, for example, electrophoretic equipment or centrifuges.

  For each safety issue identified, list in your logbook the safety controls and precautions to be taken.
- Any ethical issues associated with the conduct of your experiment, as, for example, the need to seek informed consent if your experiment will involve human subjects.

Methods: Identify your independent variable and how you will change it during the experiment. Identify the dependent variable expected to respond to these changes and identify how you will measure the changes. Identify all the controlled variables to be kept constant throughout the experiment.

Figure 17.17 shows an experimental set-up designed by a team of students to examine the following hypothesis: *IF the rate of photosynthesis in plants is dependent on the intensity of light, THEN changes in light intensity will cause a change in photosynthetic rate.* The students predicted that the direction of this change would be a fall in the rate of photosynthesis as the light intensity decreased. Check out their experimental design.

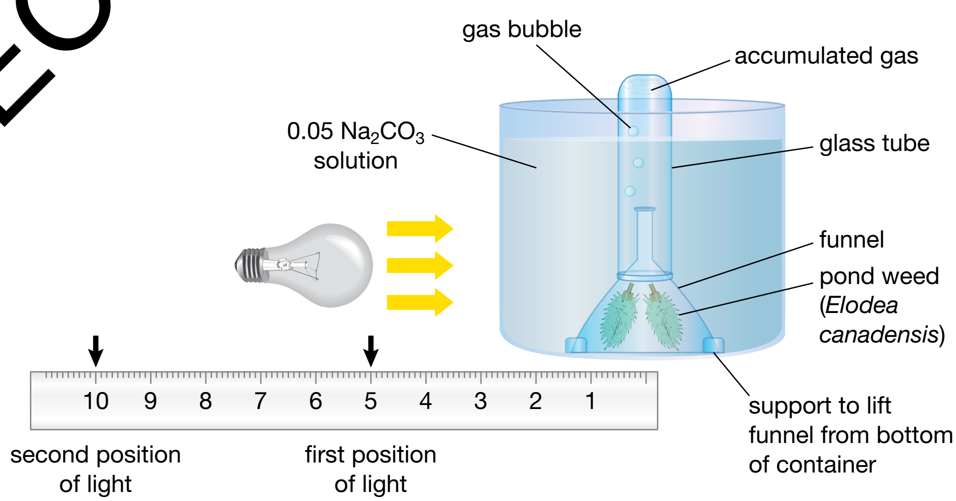


FIGURE 17.17 Diagram showing the experimental set-up used by students investigating whether a link exists between the rate of photosynthesis and the intensity of incident light.

Examine this figure and note the following:

1. The **independent variable** in this experiment is light intensity. The students varied this independent variable by progressively moving the light source along a one-metre rule. The students selected a starting point for the light source of 5 cm from the pondweed, then 10 cm, then 15 cm, then 20 cm,

after which they moved the light source in 10 cm intervals up to 50 cm away from the pondweed. This procedure gave them seven data points.

Earlier data gathering by the students led them to discover that the light intensity (I) at any point is inversely proportional to the square of the distance from the light source; that is, intensity (I) is proportional to $1/d^2$. So, the students decided to denote the light intensity at the starting point of 5 cm as 100 per cent and express the intensities at the other distances as a percentage of that value (see table 17.5 below).

2. The **dependent variable** is photosynthetic rate, but the students had to decide how this could be measured. The students considered the equation for photosynthesis and asked: *Can we easily measure the rate of disappearance of reactants or the rate of appearance of the products of this reaction?*

$$6CO_2 + 12H_2O \xrightarrow[\text{light}]{\text{chlorophyll}} C_6H_{12}O_6 + 6O_2 + 6H_2O$$

- The students recognised that measuring photosynthetic rate could be done by measuring the rate of fixation of carbon dioxide into glucose. However, the students rejected this because this method would require providing plants with a source of radioactively labelled C (as carbon dioxide) and tracking the disappearance of carbon dioxide and/or the appearance of glucose.
- Because water is both a reactant and a product in photosynthesis, tracking water poses a major problem. This method was also rejected by the students.

The students decided to measure photosynthetic rate using the production of oxygen and, to visualise the appearance of oxygen, they decided to use an aquatic plant, such as *Elodea canadensis*, the common pondweed. Pondweed is often used in freshwater aquaria, and the oxygen produced by the plant appears as bubbles released from the cut end of the pondweed. In reaching this decision, the students made a number of assumptions: (i) the bubbles of gas released from the cut stem of illuminated pondweed consist of oxygen, and (ii) all the bubbles are the same volume. Using these assumptions, the rate of production of gas bubbles from the pondweed could be identified as directly proportional to the rate of photosynthesis. This meant that the students could use gas bubble production as a valid substitute for the rate of photosynthesis in the pondweed. So, in this experiment, the dependent variable is the rate of gas bubbles produced (bubbles per minute) by the pondweed.

3. The **controlled variables** in this experiment identified by the students were as follows: temperature, carbon dioxide concentration, and background lighting. The students knew that they should control these variables so that any changes they observed in the rate of gas bubble production could be identified as due to changes in light intensity.
   - To control for temperature fluctuations, the students decided to use a 20-watt light-emitting diode (LED) as the visible light source. LED lights have a heat sink that absorbs any emitted heat. A 20-watt LED has a light output that is equivalent to that of a 100-watt incandescent light. As a back-up, they decided that if they had to use another kind of light source, the large volume of water around the funnel and test tube would act as a buffer against any significant rise in temperature.
   - To control for carbon dioxide concentration, the students decided to immerse the pondweed in 0.05 M sodium hydrogen carbonate ($NaHCO_3$) solution. When dissolved in water, sodium hydrogen bicarbonate undergoes a reaction and one of the products is carbon dioxide. One student made the important suggestion to put two small blocks under the funnel

to lift it so that the sodium hydrogen carbonate solution could freely move. If this was not done, the concentration of carbon dioxide in the funnel could have depleted.

- To control for background lighting, the students decided to set up their experiment in a darkened area to minimise any stray incident lighting. Just enough light was needed to count the number of bubbles released per minute.

To record their results, the students prepared a table as shown below. Note how the light intensity decreases with increasing distance between the light source and the centre of the funnel with the pondweed.

TABLE 17.5 The rate of photosynthesis and intensity of incident light

Distance from light source (cm)	Percentage of initial light intensity	Number of bubbles per min
5	100	
10	25	
15	11	
20	6	
30	3	
40	2	
50	1	

Control groups and experimental groups

The design of many experiments includes a **control group** as well as one or more **experimental groups**. The experimental groups are exposed to the changing conditions determined by the independent variable, but the control group is independent and not affected by those conditions.

The control group serves several purposes: it can show that the experiment is working and it can provide a baseline result against which the results of the experimental group can be compared.

In the above experiment, the control would be a set-up without pondweed. Why is this needed?

In an experiment concerned with the effect of salinity on plant growth, a control group of seedlings would be a tray of seedlings allowed to grow under conditions of zero salinity, while the experimental groups would be exposed to conditions of different salinities.

For your study, you will need to decide how many experimental groups you will need and the size of each group.

Sample size

The size of the control and experimental groups is an important factor in experimental design. The size of each group must be sufficiently large so that (1) replicate results can be obtained and a meaningful comparison made between the outcomes in the control and the experimental groups, and (2) the results from the experimental group can be seen as applicable to the larger population. The upper limit on sample size is determined by cost and space considerations and other practicalities.

Living organisms, whether they be plants, algae or animals, are an uncontrolled source of genetic variation that can interact with the variables under investigation. Sample sizes that are sufficiently large produce replicate results or observations that (3) can reduce the effects of this uncontrolled variation.

Complicated formulae exist for calculating minimum sample sizes. However, one simple rule-of-thumb is that ten observations are required for each experimental variable (see question 8 in the end-of-chapter questions).

Minimisation of bias

Bias is an intentional or unintentional influence on a research investigation as a result of systematic errors introduced by a researcher into the sampling or the testing procedures of an experiment. These biases will prejudice the research findings and raise questions about their validity and reliability.

Consider a clinical study into the effects of a new drug that involves a control group of individuals who will receive a placebo and an experimental group of individuals who will receive the drug under test. Various kinds of bias, as identified below, could affect the research findings, and various means of minimising that bias are also identified:

- **Selection bias**: This type of bias can arise unless participants are not equally and randomly assigned to the experimental and the control groups. All participants should have an equal chance of being placed in either group. Selection bias can be minimised by randomly allocating participants to each group.
- **Sampling bias**: This type of bias can arise if the participants chosen for the study are not representative of the target population. If this occurs, the research results cannot be generalised to that population. For example, findings from a study carried out using adult male subjects cannot be applied to the general population because the latter includes children and female adults. Sampling bias can be minimised by ensuring that the participants in the study are a reasonable representation of the target population.
- **Response bias**: This type of bias arises when only certain members of the target population respond to an invitation to participate in the clinical trial, resulting in an unrepresentative sample of the larger population. Response bias can be minimised by ensuring that the participants in the study are a reasonable representation of the target population.
- **Measurement bias**: This bias can arise, even unintentionally, if the people measuring the test results are aware of whether they are measuring the results of individuals in the control group or in the experimental group. There is a possibility that the people measuring the results will, even unintentionally, focus on data that supports the drug treatment. Measurement bias can be minimised ideally by designing the clinical trials as a double-blind study in which neither the participants nor the researchers know which people are in which group.

In the case of an observational study, an investigator must design a plan of action to make observations, to collect samples, perhaps over a period of time, from the area under investigation. The design will also need to identify how the analysis of the information gathered will proceed. As for an experimental study, the investigator will have a working hypothesis that will be tested.

Step 5. Conduct the experiment and record results

Your teacher will supervise your laboratory or field experiment, or your observational study, and will ensure that health and safety requirements are met. Your teacher will also ensure that your investigation is carried out according to any ethical principles that may apply.

Record your progress throughout the experiment and the results of your experiment in your logbook. Typically, if the variables in your experiment are numerical, you will set up a table and record the values of the independent variable and those of the corresponding dependent variable. If appropriate, calculate any average values.

Displaying results of investigation

Once you have your results recorded in a table, you must display your data in an appropriate form. Results can be displayed in various ways, depending on the nature of the variables and what you want to show. Commonly used representations include:

- scatter plots or scatter diagrams
- bar graphs
- line graphs
- mapping.

Presenting results in a graph makes it easy to see relationships between the variables involved because graphs provide a visual image that can highlight trends and reveal patterns. Time sequence graphs are important because they can display key turning points in a dataset. Visual representations such as diagrams and maps are very useful tools for simplifying large amounts of data, for revealing trends, for telling stories, and for communicating information. The map of cholera deaths in Soho that was drawn by John Snow made clearly visible the clustering of these deaths around the Broad Street water pump.

Scatter plots are also termed x–y graphs. **When you have paired numerical data, a scatter plot can be used to see if a relationship exists between the two variables**, as, for example, the weights of a sample of adults plotted against their heights. Each dot represents one observation, with the position of the dot determined along the x-axis and y-axis by the value of each of the variables concerned, namely weight and height (see figure 17.18). The independent variable is plotted on the x-axis and the dependent variable on the y-axis. Note that each axis has a label and measurement units. A scatter plot can show how much one variable is affected by the other. If data points fall along a line or a curve, the variables are linked or correlated, and the closer the data points are to the line, the higher the correlation. In the figure below, a strong positive correlation is displayed. What shape would you expect in a scatter plot showing ages of a sample of adults and their reaction times?

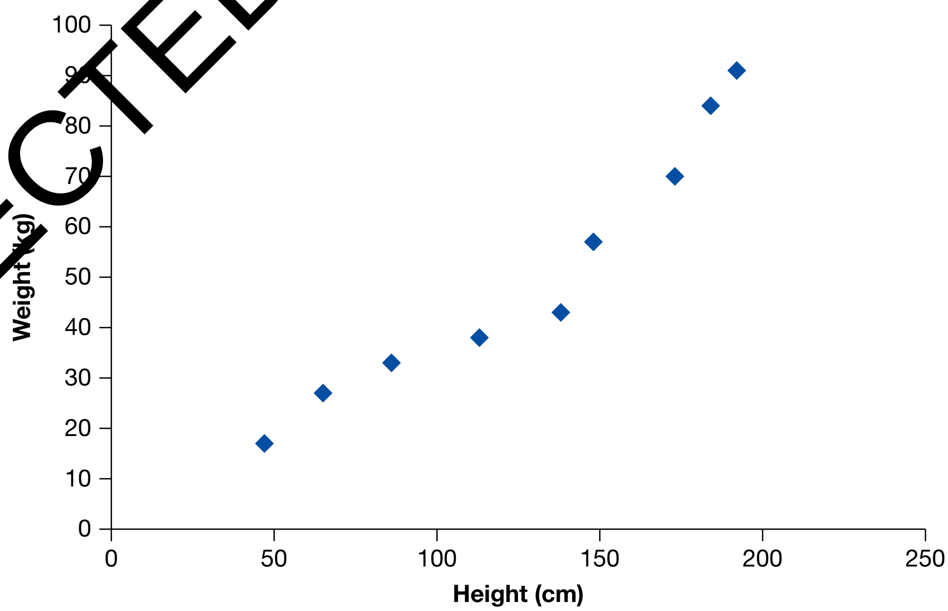


FIGURE 17.18 Scatter plot showing the relationship between the height and weight of a sample of adults.

Bar graphs can display numerical data, such as numbers or frequencies (percentages) of one or more discrete categories or groups. Figure 17.19a is a bar graph in which the vibration rates of the wings of various insects are represented by vertical bars. The horizontal or x-axis has no scale because it simply shows categories. The vertical or y-axis has a scale showing the units of measurements — in this case, vibrations per second. Note that the bars are separated by spaces. Of the insects shown, which has the greatest wing vibration rate?

Bar graphs can also be used to compare two (or more) sets of data by using side-by-side bars, as shown in figure 17.19b. This bar graph shows the percentage increase, relative to the control, in two algal species growing under conditions with added mineral nutrients. Which nutrient caused the greatest increase? Which species of alga (A or B) showed the greater increase?

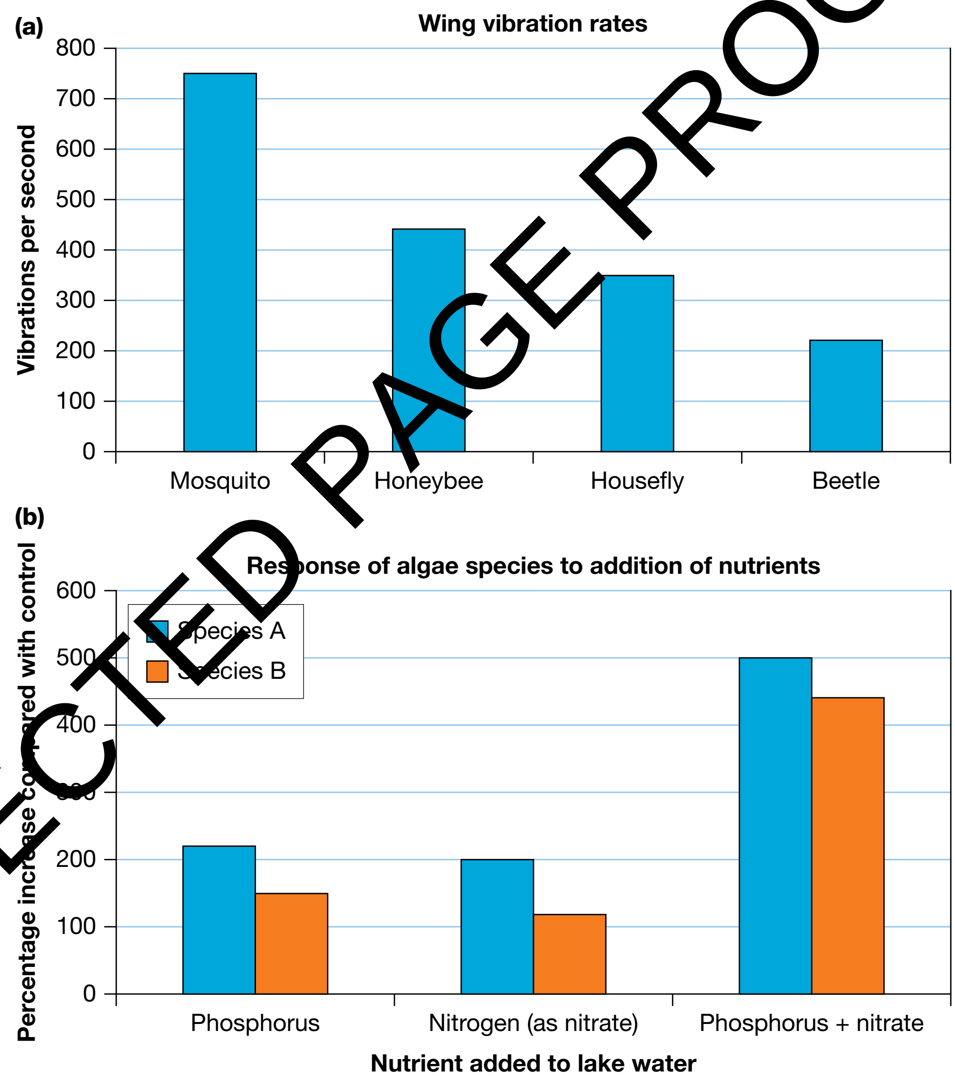


FIGURE 17.19 **(a)** Bar graph showing wing vibration rates (vibrations per second) for four insects. **(b)** Percentage increases, compared to control, of two species of algae in response to the addition of nutrients to the water in which they were growing

Less commonly, bar graphs are drawn using horizontal bars. This orientation is used when the category label is long and complex and cannot be fitted along the x-axis. For an example, refer to figure 8.10a in the text.

Check out the bar graph in figure 8.67 in the text. How many data sets are shown on this bar graph? What story does this bar graph tell? Is it easier to tell this story in words or as a bar graph?

A **histogram** is a special kind of bar graph (see figure 17.20). In a histogram, the data represent continuous categories rather than discrete values. In contrast to the situation with a bar graph, a histogram has a scale on both the x-axis and the y-axis, and the bars are not separated by spaces.

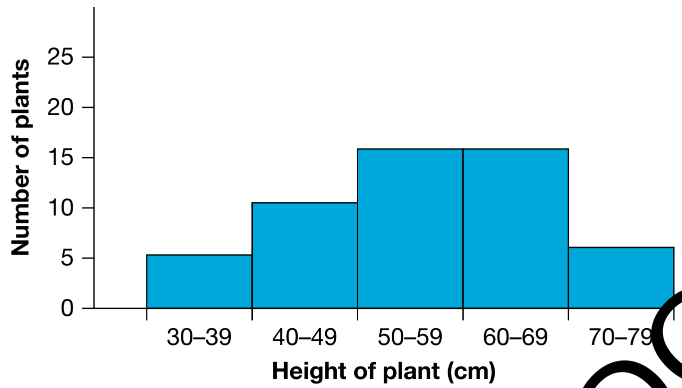


FIGURE 17.20 A simple histogram showing the results of a survey of the heights of a particular plant species growing on a moorland.

In a **line graph**, a series of dots represents the values of a variable, and the dots are joined using a straight line. Note that the ends of a line graph do not have to meet the axes of the graph. Line graphs are often used to show changes over a continuous period of time, or over space. In particular, **line graphs can show changes in a variable over time and so identify patterns, trends and turning points in a dataset.**

A line graph can be used to show the changes in a dependent variable in response to changes in an independent variable. The students investigating the relationship between light intensity and rate of bubble production showed their results as a line graph with each point showing the bubble production for a given light intensity. In this graph, the vertical axis showed the value of the dependent variable (number of bubbles per minute) and the horizontal axis showed the percentage light intensity.

Data on two variables can be plotted on the same line graph to show the change in two variables over the same time period. For example, refer to figure 8.81 in the text, which shows the variations in the population size of a prey species (moose) and its predator species (wolves) over a three-year period. What pattern is apparent?

Step 6. Analyse results and draw conclusions

Firstly, you should review your data and ask: Is the data set complete? Do you need to collect more data?

The results of any experiment will produce data that you will need to analyse. Aspects of data that you may need to consider include the following: validity, precision, accuracy, reliability, limitations of data and sources of error.

Analysing data

1. **Validity of data**: The validity of data refers to the credibility of the research results from experiments or from observations. Validity refers to both the research findings (so-called internal validity) and to the applicability of the results to other situations (so-called external validity).
   - **Internal validity refers to the degree to which the experimental procedures measure what they are supposed to measure.** Testing internal validity asks questions such as: Can the results be believed? Could another unknown variable have influenced the results? Could any cause and effect relationships identified be explained by other factors? Internal validity is favoured when an experiment is carefully designed and a scientific approach is used.

- **External validity refers to the extent to which research findings can be generalised to the greater population.** Testing external validity asks the questions: Is the sample of the population that was used in the research study representative of the greater population? Can we be reasonably sure that the results of the research are applicable to the greater population? External validity is favoured by an experimental design that includes the use of a control group, having control and experimental groups of sufficient sizes, and randomly assigning subjects to the control and test groups.

2. **Precision**: How close to one another are repeated measures of the same variable under the same conditions? If the measurements are close, as, for example, 19.5, 19.4, 19.6 and 19.5, they are precise. **Precision refers to the closeness of repeated measures or observations even if they are not accurate.** Precise values will differ from one another because of random observational error.
3. **Accuracy**: Accuracy is different from precision. **Accuracy refers to the closeness of an experimental measurement to an accepted or known value.** Accuracy in a research investigation relies on selecting the most appropriate instruments for any measuring tasks and using them correctly.
   Accuracy is an issue in a confirmatory experiment where the aim is to confirm a known value, as, for example, identifying the diploid number of the yellow fever mosquito, *Aedes aegypti*. If you concluded from your microscopic observations that the diploid number was six, then your result would be inaccurate. (The known number is 2n = 8.) However, when an experiment is an exploratory investigation rather than a confirmatory exercise, matching an accepted value of a variable is not an issue.
4. **Reliability**: If your experiment were to be repeated a second and third time, would the results be the same? **If the same results are obtained in repeats of an experiment, the data can be said to be reliable.** Replication reduces variability in experimental results, increasing their significance and the confidence level with which a researcher can draw conclusions. Ideally, replications of experiments are carried out not only by the original investigators but, importantly, by third parties who use the same materials and the same methods as those of the original investigators.

FIGURE 17.21 Repeating a biology experiment is a means of assessing the reliability of experimental findings.

Replication is an important means of verifying experimental results and showing that they are reliable. In January 2014, a young Japanese researcher, Haruko Obokata, published in the leading science journal *Nature* an article that was hailed as a major breakthrough in stem cell research. Obokata had apparently developed a very simple method of producing pluripotent stem cells from adult somatic cells — just soak them for 30 minutes in a weak solution of citric acid. However, in spite of the simplicity of her method, no one else was able to replicate her results. In reality, Obokata had falsified and fabricated her experimental data. The fact that her results were not reproducible was critical and resulted in her retracting her research findings.

5. **Limitations of data**: The data that is gathered from the experimental results will have limitations. Limitations arise from several sources that can affect the quality of the data:
   - Experiments create artificial situations that do not necessarily represent real-life situations.
   - While every effort may be made to identify controlled variables and keep them constant throughout the course of an experiment, it is not always possible to identify and control every one of this type of variable. Therefore, the result obtained from an experiment may be confounded by the influence of an uncontrolled variable.
   - The degree on which results obtained in the laboratory can be generalised to other situations and applied in the real world is limited.

Limitations in models used in biology

As well as data limitations, **limitations exist in the models that we use in biology.** Models can be physical models, or mathematical models or conceptual models. Models can provide an explanatory framework that explains observed phenomena and helps organise myriad data into a related whole. Physical models can make an abstract concept tangible.

While they are very useful, models have their limitations.

- Because of the complexity of the processes that operate at every level of biological organisation, models cannot include all the details of the processes or the things that they represent.
- Models are necessarily approximations of the real world; for example, one model of population growth is the S curve of density-dependent growth (refer to chapter 8 in the text). This mathematical model helps predict how populations might grow over several generations; however, this model is an ideal simplification since it cannot take into account all the environmental variables that can affect populations, and so its predictions can differ from what happens in real-world populations.
- In making a concept or object concrete, a model must compromise accuracy; for example, figure 17.22 shows various molecular models of ATP, the energy supplying molecule of cells. Note, at the top of the figure, ATP is shown simply as its structural formula with chemical symbols showing the arrangement of its atoms and the bonds between them. In the middle, ATP is shown as a ball-and-stick model with atoms shown as coloured balls in their 3D positions joined by sticks that represent covalent bonds. The colour of each ball represents a particular chemical element, with carbon atoms shown as black balls, oxygen atoms by red balls, and hydrogen atoms by white balls. At the bottom, ATP is shown as a space-filling model in which the radius of each ball is proportional to the radius of the atom concerned.

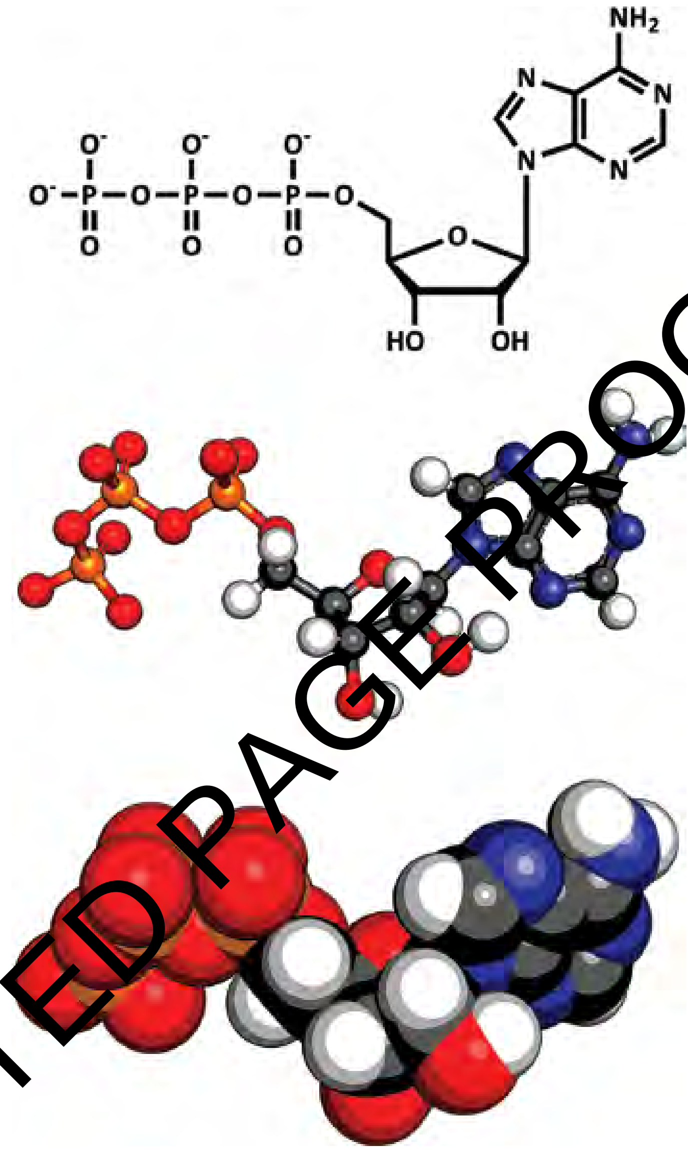


FIGURE 17.22 Molecular models of ATP, the chemical energy source of cells. At the top, the structural formula of ATP is shown. At the centre, ATP is shown as a ball-and-stick model. Which atoms are denoted by blue balls? By orange balls? At the bottom, ATP is shown as a space-filling model. Do all models show the bonds between the atoms?

Sources of error in investigations

Several sources of error can be identified in an investigation.

- **Systematic errors** produce measurements that are consistently too high or too low, for example:
  - instrumental errors that arise because an instrument, such as a weighing balance for which the zero has not been reset before use
  - environmental errors that arise because of malfunctions, such as a power outage, that affect the conditions under which an experiment is being conducted.
- **Random errors** are chance variations in measurements that are equally likely to be high as low. An example of a random error is an error of judgement when reading the smallest division on the scale of a measuring instrument, such as a micrometer, a burette or a measuring balance. However, if multiple observations are made and an average calculated, these random errors are minimised.

It should be noted that so-called human 'errors' are not errors but are mistakes that result from carelessness. Examples include gross misreading of an instrument, and writing the wrong result in your logbook, such as 40 instead of 4.0.

Hypothesis evaluation and conclusions

Evaluating a hypothesis uses the evidence from the results of the experiment that was designed to test the prediction from that hypothesis. In John Snow's investigation, his prediction was that stopping the supply of water from the Broad Street pump would reduce the incidence of cholera.

If the prediction from your hypothesis was validated by your experimental results, you should evaluate your hypothesis as 'supported' and your tentative explanation of the scientific observation now stands as a reasonable explanation. If your predicted outcome was not supported by your experimental results, you should evaluate your hypothesis as 'refuted' or 'disproven.' In that case, you might think about how you could modify your hypothesis.

Other conclusions might relate to the following questions.

- Were there any outliers in your dataset? An **outlier** apparently doesn't fit — it is a data point that is distant from all the other data points.
  If you had an outlier, what did you do about it? Rather than ignoring it, you should try to account for it. Most commonly, it will either be a systematic error or an outright mistake in measurement or in calculation, and so can be dropped. Occasionally, an outlier can be a legitimate observation that warrants further investigation.
- What future improvements can you suggest regarding your experimental design?
- In light of your experimental results, what follow-up experiments might you suggest?
- How do your research findings link to the key biological concepts relevant to your research question?

Step 7. Communicating findings

The results of a scientific investigation are typically communicated to others. In your case, you will be required to communicate the findings of your experimental or field study, or your observational study. The curriculum design identifies that this may be done using a scientific poster, a practical report, an oral communication or a digital presentation. In presenting your findings, ensure that you use correct biological terminology and relate your findings to relevant biological concepts.

When you present a conclusion, ensure that you provide the supporting evidence. However, evidence can be strong or weak. Check out table 17.6, which summarises the contrasting features of strong and weak evidence.

TABLE 17.6 Table showing the contrasting features of strong and weak evidence.

Strong evidence	Weak evidence
• is based on facts derived from studies with high internal validity	• is based on personal opinion or on anecdote
• is related to the topic	• is irrelevant or unrelated to the topic
• provides support for stated conclusions	• provides little support for stated conclusions
• is logically organised	• is not logically organised
• is free of bias	• may contain bias
• comes from a source that is trustworthy and reputable	• comes from a source that is unknown or untested

Remember: If you use any material that is the work of another person, you *must* acknowledge the material's source. Do not claim it as your own work. Acknowledgments come in two formats: a short version when it occurs in the body of a report or a poster, and a longer version when it occurs in the *Reference and Acknowledgments* section at the end of a scientific report or poster. These two formats are described in the box at the end of this section.

If you quote the exact words of one or more sentences from a source, ensure that the quotation is enclosed within inverted commas and that the source is acknowledged at the end of the quotation. This is not necessary if you express in your own words the ideas taken from another source.

Communication in science may be done through a variety of avenues such as publications in scientific journals, science reports on TV (refer to figure 1.39 in the text), articles in popular science magazines such as *New Scientist* and *Scientific American,* newspaper articles, books, brochures from health organisations, and web-based publications, such as Web MD. The language used for communication should reflect the intended audience. A research report in a scientific journal uses language that is very different from a newspaper account of the same research.

However, communication does not guarantee immediate acceptance of the findings of a scientific investigation. In the case of the Australian researchers Barry Marshall and Robyn Warren, their research into the cause of peptic ulcer disease and gastritis eventually led to a complete revolution in the treatment of these conditions. By demonstrating that peptic ulcer disease was the result of infection of the stomach lining by *Helicobacter pylori* bacteria, Marshall and Warren recognised that this condition could be completely cured using a short course of antibiotics and stomach acid inhibitors.

At the time, these were chronic and often disabling conditions, believed to be the result of stress and lifestyle. Sufferers required lifelong treatment and, in some cases, even drastic steps such as surgical removal of part of the stomach. Marshall and Warren published a detailed summary of their research findings in 1985. These findings were initially dismissed by specialists in the field (gastroenterologists). By the mid 1990s, however, the research findings of Marshall and Warren had become mainstream, and the treatment of peptic ulcer disease and gastritis by a short course of antibiotics was routine. The ultimate vindication of Marshall and Warren's research came in 2005 when they were co-recipients of the 2005 Nobel Prize in Physiology or Medicine.

REFERENCING DATA SOURCES

The examples below are based on the *Harvard Style Guide* published by Monash University in 2012; available at the library website of Monash University.

For each type of publication, the first style (A) gives full details and is used in reference lists at the end of manuscripts or posters. The second style (B) is short and is used in the body of the text of a manuscript or a poster.

- **Online authored articles**:
  Style A: Author name and initial/s (Year of publication) Title of document. Retrieved on [date] from http:// URL
  Example: **Geggel L (2016) Snakes used to have legs and arms . . . until these mutations happened. Retrieved on 4 Nov 2016 from www.livescience.com/56573-mutation-caused-snakes-to-lose-legs.html**
  If the online article is undated, put (n.d.) in place of the date
  If there is no author, use the first line of the text of the article in place of the author's name
  If there are up to three authors, list them all.
  If there are more than three authors, only use the name of the first author and follow it by the phrase 'et al' (meaning and others).
  Style B: Example: **(Geggle 2016)**
- **Books**
  Style A: Author name and initial/s (Year of publication) *Book title*. (Edition if applicable.) Place of publication, publisher. Page number(s) you accessed.
  Example: **Langley P & Jones M (2013) The search for Richard III: The king's grave. London, John Murray.**

If there are more than three authors, only use the name of the first author and follow it by the phrase 'et al.'

Style B: Example: **(Langley & Jones 2013)**

- **Journal articles**:

  Style A: Author name (year of publication) *Title of article*. Title of journal. Volume number. Page numbers.

  Example: **Smith SL (2007) *Stature estimation of 3–10 year children from long bone lengths.* J Forensic Sci vol. 52 pp. 538–546.**

If there are more than three authors, only use the name of the first author and follow it by the phrase 'et al.'

Style B: Example: **(Smith 2007)**

- **TV programs:**

  Style A: Title of program (date) (TV program), channel identification.

  Example: **Gene editing made simple 2016, television program, ABC Sydney 30 August**

  Style B: Example: **(Gene editing 2016)**

KEY IDEAS

- Designing an experiment means constructing a detailed experimental plan to test a hypothesis in advance of doing the experiment.
- Control and experimental groups are typically incorporated into the design of an experiment.
- Bias is an intentional or unintentional influence on a research investigation as a result of systematic errors introduced by a researcher into the sampling or the testing procedures of an experiment.
- Results of investigations may be presented using visual representations, including graphs, diagrams and maps.
- Data may be assessed in terms of attributes, including validity, accuracy, precision, and reliability.
- Evaluation of hypotheses results in hypotheses being either supported or refuted, but never proven.
- Communication of the findings from scientific investigations is an important step in a scientific approach.

QUICK CHECK

6 What is a double-blind study?
7 What kind of bias is minimised by using a double-blind study?
8 Identify one advantage of showing the numerical values of two variables in a scatter plot, rather than in a table.
9 A student group gathered data on the abundance of several species along a transect line. Suggest how this data set might be displayed. (If you want a clue, check out figure 8.67 in the text.)
10 What is the difference between precision and accuracy in relation to data?
11 How does a systematic error differ from a random error in an investigation?
12 Identify four channels of communication for the findings of scientific investigations.

Chapter review

studyon

Sit exam

Key words

accuracy
bar graphs
categorical variables
control group
controlled variables
dependent variable
experimental groups
falsifiable
germ theory of disease
histogram
independent variable
limitations of data
line graph
measurement bias
miasma theory of disease
numerical variables
outlier
precision
random errors
reliability
response bias
sampling bias
scatter plots
selection bias
systematic errors
testable
validity of data

Questions

1 **a** Identify three features of a scientific hypothesis.
In medieval times, 'explanations' of the cause of the plague included divine punishment for sin.
b Is it possible to formulate a hypothesis to test this explanation? Explain.

2 **Identify the following statements as true or false:** →
a Interpretation follows data collection.
b The dependent variable in an experiment must be kept constant throughout the course of an experiment.
c If your experimental result causes you to reject the hypothesis under test, this means that you made a mistake in your experiment.
d A scientific investigation starts by carrying out an experiment.

3 **You conduct an experiment and record the data. You notice some of the data doesn't fit with what you expected or doesn't agree with the other data. Which of the following actions should you take?** →
A Toss it out because it looks wrong.
B Figure out what you want the data to show and then decide if you will keep it.
C Try to identify why this occurred, then decide if it should be kept or discarded, and discuss it in your results analysis.

4 **Which of the following is a requirement for a good hypothesis?** →
A You need to make the hypothesis broad so that the data can support it or reject it.
B You need to be able to design and conduct an experiment that can test the hypothesis.
C You need to know in advance whether or not the hypothesis is true.

5 **Examine figure 17.23.** →
a What data is shown by the blue bars?
b What is represented by the black line graph?
c In which two months did the rainfall show the greatest deviation below the long-term average?
d In which two months did the rainfall show the greatest excess above the long-term average?

6 **Identify a key difference between the members of the following pairs.** →
a independent and dependent variables
b qualitative and quantitative data
c control group and experimental group
d primary and secondary sources of data
e categorical variable and numerical variable
f an experimental study and an observational study
g data and conclusions

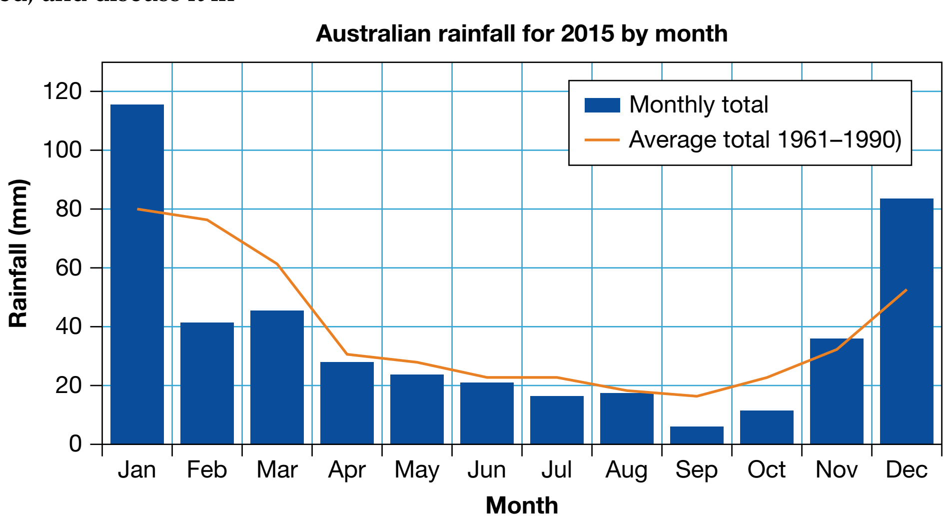


FIGURE 17.23 Australian rainfall for 2015 by month

7 Consider an apple. ➜

a Identify pieces of three quantitative data that you could observe.

b What instruments, if any, would you need to make these observations?

c Identify three pieces of qualitative data that you could observe.

d What instruments, if any, would you need to make these observations?

8 You wish to see if a new drug is an effective treatment for asthma. One hundred volunteers are randomly divided into two groups (A) and (B). Members of group B are given the drug and members of group A are given a placebo in the form of a sugar pill. ➜

a Which is your test group and which is your control group?

b This experiment was carried out as a double-blind study.

i What does this mean?

ii What is the advantage of this procedure?

c A student said, 'Wouldn't it be better to use all 100 volunteers in a larger test group, rather than split the volunteers into a test group of 50 and a control group of 50?' How would you respond to that student?

9 A joint report by the Australian and Queensland governments details the health of the Great Barrier Reef using a variety of measures, one of which is the concentration of herbicides that are inhibitors of a key stage of photosynthesis (PSII). Figure 17.24 shows the maximum concentrations of these herbicides in the Mackay Whitsundays region of the Great Barrier Reef from 2008–09 to 2014–15. ➜

a Suggest one or more possible sources for these herbicides.

b What would be the expected impact of these herbicides on the seagrasses in these areas? Explain your answer.

c How might high levels of herbicides affect the ecology of the reef?

The different colours of the bars denote the categories of herbicide in terms of their impact on photosynthesis, with category 5 having no impact and category 1 having the greatest impact on photosynthesis.

d Where and in which years did the highest and most damaging herbicide concentration occur?

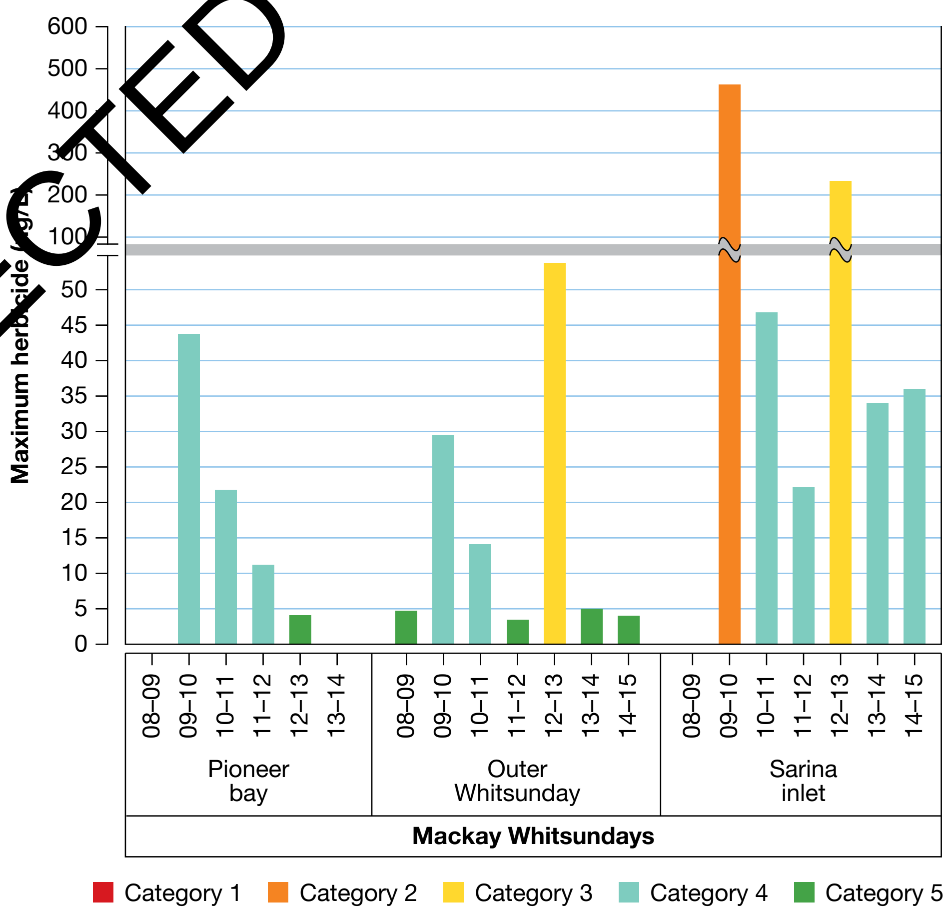


FIGURE 17.24 Bar graph showing the maximum concentrations of herbicides that inhibit photosynthesis in the period from 2008–09 to 2014–15 in three regions of the Mackay Whitsundays region of the Great Barrier Reef. Note the break and the change of scale in the vertical axis. (Source of data: Great Barrier Reef Report Card 2015: Marine results page 14.)

10 Consider a simple experiment to investigate the relationship between wavelength of light and growth rate in which trays of seedlings are grown under conditions with different wavelengths. ➜

a Identify the independent variable in this experiment.
b Identify the dependent variable.
c Identify at least four controlled variables that should be identified by the investigator.
d Which kind of variable is:
  i kept constant throughout the experiment
  ii manipulated by the investigator during the experiment
  iii observed during the experiment and changes recorded?
e Would a sample size of two seedlings, one test and one control, be a good design? Explain.

11 Consider the following data: ➜

  i labelling of S, M, L, XL, XXL on items of clothing
  ii labelling of 300 mL, 600 mL, 1 litre, 2 litre on drink cartons
  iii values of 73 °C, 86 °C, 54 °C, and 92 °C identified as maximum daily temperatures
  iv classification of cats as Birman, Siamese, Tonkinese, Manx, Russian Blue.
a Which, if any, are examples of a categorical variable?
b Which, if any, are examples of a numerical variable?

12 Examine the graph titled *Phytoplankton and Zebra Mussels in the Hudson River* shown at the Cary Institute weblink and answer the following questions: ➜

a What kind of graph is this?
b What variable is represented by the blue data points?
c What variable is represented by the red data points?
d Phytoplankton are microscopic photosynthetic organisms found in bodies of freshwater. What was the peak concentration of phytoplankton in the Hudson River?
e The concentration of chlorophyll a has been used as a substitute for the concentration of phytoplankton. Can you suggest why this was done?
f What trend is apparent in the phytoplankton population over the period shown?
g Zebra mussels (*Dreissena polymorpha*) are filter feeders, and this non-native species reached the Hudson River in 1991. What happened to the phytoplankton population in the period after 1991?
h Write a hypothesis that might explain the sudden and persistent decline in the phytoplankton population.
i Suggest how this hypothesis might be tested.

eBook*plus*
Weblink
Cary Institute

13 Consider the hypothesis: *IF the growth rate of plants depends on the duration of sunlight they receive, and seedlings are exposed to varying periods of light, THEN seedlings exposed to longer periods of light will grow faster than those exposed to shorter periods.* ➜

a Identify the independent variable.
b Identify the dependent variable.
c State the prediction that would be used to test this hypothesis.
d Suggest a possible means by which the growth rates might be measured.
e Identify three controlled variables that would need to be taken into account when designing an experiment to test this hypothesis.

14 Consider the following table that shows the results of an experiment that looked at the effects of salinity on plant growth: ➜

a How many replicates were used in the control group? In each experimental group?
b Why not use just two or three?
c What is the independent variable in this experiment?
d What is the dependent variable?
e Identify three possible controlled variables.

TABLE 17.7 The effects of salinity on plant growth.

Concentration of salt water (mg/L)	Replicates: seedling root length (mm)										Average length
	1	2	3	4	5	6	7	8	9	10	
0 (control)	20	16	16	9	14	13	13	10	12	10	13.3
0.75	14	22	13	14	15	16	11	16	17	10	14.8
1.5	15	19	12	18	10	7	0	9	4	13	11.8
3	9	7	6	6	16	11	10	17	11	6	9.9
6	1	1	0	0	0	0	0	0	0	0	0.2
12	1	0	0	0	0	0	0	0	0	0	0.1