IV

Evaluating the evidence
Thinking straight about the world is a precious and difficult process that must be carefully nurtured.

Thomas Gilovich

Whether sick or not, we are continuously offered health advice. When we are sick, we are told that a new treatment will help us. When well, we are told how to avoid getting sick. Sometimes we’re encouraged to have – or perhaps to avoid having – a screening test to detect disease early. Often we are warned that something causes cancer, heart disease, high blood pressure, and so on.

Some of this advice may be sound, but much is not. The important question is ‘How do you tell which is which?’ The answer is simple – ask if there is valid evidence to support the advice, irrespective of how competent or qualified the person giving you the advice seems to be. Competent practitioners should be prepared and willing to discuss the evidence supporting any advice that they offer.

The best way for you to decide whether the advice is good is to find out how good the supporting evidence is. The amount of effort that you put into this depends upon the importance of the decision. For minor issues you may not want to spend too much time, but, if it is an important decision for you, it will be well worth your while exploring the reliability of the evidence.

To do this, you need to find and make sense of the research literature on which the evidence is based. Don’t be daunted. This is
Chapter 10

not nearly as complicated as it may sound, and doesn’t require years of medical training. What it does require is the ability to distinguish between research that is well done and free of biases and research that is shaky and subject to biases and misinterpretation. This may sound difficult but is surprisingly straightforward if you follow some basic guides described in the following pages.

You can apply these guides to information in the media or on websites or to decisions made with your practitioner when you feel that the situation requires it. They can also be applied to articles in the medical literature, abstracts of which are available on the internet, as described in Chapter 15. Abstracts of studies are often structured with headings compatible with the following guides.

Is there an evidence-based practice guideline or systematic review of randomised trials?

As discussed in Part II, when you are examining the evidence about an intervention, the first question should always be: Is there an evidence-based practice guideline or systematic review of randomised trials? A guideline or systematic review based on the most recent, valid research, recommending how best to manage the disease or condition, is the single best source of evidence that you can get.

Be aware, however, that we are recommending evidence-based guidelines and reviews that come from the accumulated high-quality evidence from studies, and how to apply the evidence to individuals who vary in their preferences and characteristics. Not all practice guidelines are based on strong evidence. Some are just a consensus view of a group of experts without the search for impartial evidence. This is why you should establish whether a guideline is evidence based. Either ask your practitioner or read it yourself to determine whether it refers to up-to-date, valid research as its basis. Some guidelines are specifically written for consumers. Some summaries of systematic reviews are now available that have been written for the general public. Those that are not specifically intended for consumers may require your practitioner’s help to interpret.
Remember, if a guideline or systematic review is evidence based:

- it will be recent, e.g. within the last 5 years; if it is not, ask if there is a more recent one.
- it will describe all the treatment options; systematic reviews should describe clearly what question they are addressing.
- it will describe outcomes that are person-centred (about quality of life and survival).
- it will describe both the benefits and harms of each option.
- it will describe how the best evidence was selected and report the highest level of evidence for each recommendation.

If there is no evidence-based guideline or systematic review on the health topic in question, you and your practitioner may need to make a decision based on existing studies on the disease and available tests or treatment options. We deal with a study’s relevance first because, if it is not relevant to your situation, there is no point in assessing its quality.

**Is the evidence relevant?**

- The evidence should describe outcomes that are person-centred. They should tell you about survival or quality of life rather than surrogate measures such as laboratory results.
- The evidence should describe both harms and benefits and tell you how likely they are.
- The evidence should describe how tests and treatments compare with a wait and watch approach or each other.

**Are the outcomes person-centred?**

You need to make sure that a study that you are evaluating is relevant to your needs. It is important to know what the intervention’s outcomes are, when they might occur, and how permanent and probable they might be. The evidence should demonstrate an effect on the length or quality of life rather than on some substitute
intermediate endpoint. A study showing that drugs to dilate blood vessels (vasodilators) improve blood flow through the heart in people with heart failure is not, in itself, sufficient evidence for using the drugs. What would be relevant, however, is evidence that they help to reduce the symptoms and reduce the probability of requiring hospital treatment.

Are the harms described as well as the benefits?

You need to know about the chance of side effects and other downsides to treatments and tests, as well as when they might occur and if they are likely to be permanent. For example, chemotherapy for cancer may cause nausea, hair loss and weakness during the time it is taken. Another example is screening for early detection of disease. Screening may result in more people being recalled for further testing and invasive investigations than have the disease being screened for.

You also need to know the likelihood that you will benefit from treatment and tests. For example, as discussed in Chapter 9, the use of hormone therapy around the menopause may reduce symptoms of hot flushing and night sweats, but these benefits must be weighed up against its potential to increase the risk of breast cancer, blood clots, strokes and probably heart attacks.

Is the test or treatment compared with other suitable options?

Studies on new tests or treatments should show comparisons with the appropriate options – either a placebo or the best available existing treatments. Sometimes it will be important to know whether a test or treatment is better than doing nothing. But often there is already an effective treatment that is known to be superior to placebo. In such a situation we want evidence that the new intervention is going to have greater benefits and/or fewer harms than the existing one.
It has long been known, for example, that paracetamol offers effective long-term relief from the symptoms of osteoarthritis and has minimal side effects. On the other hand, non-steroidal anti-inflammatory drugs (NSAIDs) are associated with a high rate of gastrointestinal side effects. A systematic review of 15 randomised trials in the Cochrane Library showed that NSAIDs are more effective in reducing rest and movement-related pain in osteoarthritis, but the reduction in pain is small to modest (around 20 per cent). As discussed earlier, if your level of pain is higher to start with, you will probably get a greater benefit from NSAIDs. However, NSAIDs are one and a half times more likely then paracetamol to give you gastric side effects such as nausea, indigestion and even bleeding. The review reports that 19 per cent of people on NSAIDs reported gastric side effects compared with 13 per cent on paracetamol. So neither treatment is risk free, but the chance of problems with NSAIDs is a little higher and this needs to be balanced against the amount of pain relief required. In view of the side effects of NSAIDs, it seems prudent to try paracetamol first and switch to NSAIDs only if pain relief is unsatisfactory.

**Is the evidence reliable?**

Reliability of evidence about an intervention, from the most to the least reliable, would go:

1. A systematic review of randomised controlled trials
2. A randomised controlled trial
3. Non-randomised studies:
   - cohort studies and non-randomised trials
   - population-based case–control studies
   - hospital-based case–control studies
   - other study types
4. Case reports and opinions.

Earlier in this book we introduced you to randomised controlled trials and systematic reviews, but this section will take you a bit further now that you know more about making sense of health advice.
Randomised controlled trials

We will start with an explanation of randomised controlled trials (RCTs) because it is easier to understand systematic reviews if you know something about RCTs first.

Unlike the scene in the cartoon, a good trial will:

- have a control group receiving the best existing treatment (or placebo, if there is no treatment)
- randomly allocate people to the intervention and control groups
- keep practitioners and study participants blind (or masked) to who is in which group
- follow up on everyone who was randomised to the various groups at the start of the trial.

This is the type of study that showed that an effective treatment for stomach ulcers is antibiotics and raised the alarm that antiarrhythmia drugs after a heart attack might have dangerous and previously unsuspected side effects. Randomised controlled trials also quashed the long-held theory that antioxidant vitamin tablets prevent cancer.
Until RCTs showed otherwise, many doctors believed that the most effective way to treat early breast cancer was a mastectomy to remove all of the breast. They believed this because it was what they had been taught at medical school and it was what most specialists advised. This widely held belief has since been disproved by RCTs showing that women are just as likely to survive the cancer if they have less invasive surgery combined with radiotherapy.

Why ‘controlled’?

As their name suggests, randomised controlled trials involve randomly allocating patients to either the active treatment or a comparison (control) group. This provides the all-important comparison. It is not much good knowing that a treatment leads to a 50 per cent recovery rate if you don’t know how this compares with alternative treatments or even no treatment at all.

Similarly, if some people die while receiving a particular treatment, we don’t know whether these deaths are a result of the treatment or the disease. But if there is another group not taking this treatment, a comparison of the two groups could help to establish whether the death rate is significantly higher in the group being treated.

Whether we can draw conclusions from comparing the outcomes of two groups depends on them being similar. If one group is sicker than the other, for example, this may bias the result of any comparison.

Why ‘randomised’?

Randomly allocating patients to the comparison groups aims to reduce the chance of such biases occurring. It means that the groups start out with an equal chance of events occurring during the study, whether disease recurrence, side effects from treatment or symptom relief. In other words, it increases the likelihood that any differences in outcome between the groups are caused by the test or treatment and not other factors.

Sometimes randomisation is done as a cross-over trial. This means that people are first randomised to treatment A or B and then, after the outcome has been assessed, the groups cross over so that those who were receiving treatment A switch to treatment B and...
vice versa, and the outcomes are again measured. Of course, this type of trial can be done only if the outcomes are short and reversible – for example, to measure an intervention’s effect on pain relief.

It is not enough to know that a treatment has been subjected to clinical trials. This means nothing more than that it has been tried. Clinical trials do not always include a control group and participants are not always randomly allocated. And controlled clinical trials may not be randomised. You really want to know that it is a randomised controlled trial.

The importance of ‘blinding’

Blinding or masking is another important method for eliminating bias from RCTs when measuring the effects of an intervention. It aims to ensure that the researchers and participants do not know who is receiving treatment or placebo. This helps distinguish between real change resulting from the intervention and imagined change caused by the influence of enthusiasm and attention.

In an intervention study, the researchers and practitioners involved often have some prior expectation of the effect of the intervention being studied – otherwise, they probably would not be doing the study! These expectations may influence the way in which practitioners measure and record patients’ responses, and can affect the validity of the results.

In the same way, participants often have some expectation of how the treatment will affect them, and might respond to the placebo effect. As well, when we are the target of special attention – as in being part of a trial – we tend to respond in a way that can affect our feeling of well-being. This phenomenon is sometimes referred to as the Hawthorne effect and can have a great impact on any therapy that we might be receiving.

To avoid these biases, ideally everyone involved in the trial – patients and practitioners (or researchers) – should be blinded to the treatment status. When the patients and the practitioners are blinded, this is referred to as a ‘double-blinded’ study.

If you are reviewing several RCTs on a particular intervention that show differing results, you should use blinding as a deciding
factor. A study that is blinded or double-blinded is generally more reliable than the same type of study that is not.

If you want to know whether a study is an RCT and of good quality, check that the participants were randomised to the groups and that the participants and the practitioners were blinded. If these criteria are not declared, assume that they were not fulfilled.

Also check that all those randomised were followed up and included in the study results. Loss to follow-up is selection bias and may distort the comparison of the intervention and control arms.

**Systematic reviews of randomised controlled trials**

For a review to be systematic, the authors should declare whether they have:

- identified all the RCTs on the topic
- included all RCTs on the topic unless they do not meet the criteria for high-quality RCTs
- pooled the results into a large analysis, called a meta-analysis
- assessed whether there is variability in effects in different sorts of patients.

Even well-designed randomised trials can produce apparently differing results. The best method for evaluating such studies is through a systematic review, which examines the evidence from all of the good studies that have been done.

It was a systematic review, for instance, that finally accumulated enough evidence to persuade practitioners that there is value in chemotherapy for early breast cancer and antenatal steroids for infants born before term.

But be warned! Not all reviews are systematic reviews. Many reviews published in journals and elsewhere are no more than a collection of opinions that support a particular viewpoint.

The hazard of a haphazard review is obvious. Probably all of us are prejudiced and tend to focus on what we like to see. And, even worse, some of us tend to dismiss anything that does not suit our purpose. This makes it worthwhile to set certain rules before
Chapter 10

starting a review process. Reviews are scientific enquiries and they need a clear design to preclude bias.

A review that does not have a clear design to preclude bias is clearly not as reliable as a systematic review that, as its name suggests, takes a systematic approach to identifying the valid studies conducted on one topic and analysing the combined results.

Again, to determine whether a review is systematic, check whether the authors have declared that:

- all the RCTs on the topic have been identified
- all the RCTs have been included – unless they do not meet the criteria for high-quality RCTs
- the results have been pooled into a large analysis (meta-analysis).

The Cochrane Collaboration

The Cochrane Collaboration is an international movement of thousands of professionals and consumers who produce a regularly updated electronic library of the best available evidence about the effects of interventions. The Cochrane Library gives global access, to practitioners as well as consumers, to the most recent systematic reviews of RCTs on a rapidly expanding range of health topics. It contains thousands of systematic reviews of RCTs on a wide range of health treatment options. The abstracts or summaries of the reviews are available globally free of charge. In addition, some countries provide free access to the whole library: www.thecochranelibrary.com

The Cochrane Collaboration logo (opposite) illustrates a systematic review of data from seven RCTs. Each horizontal line represents the results of one trial (the shorter the line, the more certain the result) and the diamond represents their combined results. The vertical line indicates the position around which the horizontal lines would cluster if the two treatments compared in the trials had similar effects; if a horizontal line touches the vertical line, it means that that particular trial found no clear difference between the treatments. The position of the diamond to the left of the vertical line indicates that the treatment studied is beneficial.
This diagram shows the results of a systematic review of RCTs of a short, inexpensive course of a corticosteroid given to women expected to give birth prematurely. The first of these RCTs was reported in 1972. The diagram summarises the evidence that would have been revealed had the available RCTs been reviewed systematically a decade later; it indicates strongly that corticosteroids reduce the risk of babies dying from the complications of immaturity. By 1991, seven more trials had been reported, and the picture in the logo had become still stronger. This treatment reduces the odds of the babies dying from the complications of immaturity by 30–50 per cent.

As no systematic review of these trials had been published until 1989, most obstetricians had not realised that the treatment was so effective. As a result, tens of thousands of premature babies have probably suffered and died unnecessarily (and cost healthcare services more than was necessary). This is just one of many examples of the human costs resulting from failure to perform systematic, up-to-date reviews of RCTs of health care.

**Non-randomised studies**

**Cohort studies**

Cohort studies are observational studies in which people are exposed to some factor of interest (for example, diet, smoking, occupation)
at the start of the study and then followed over a time period sufficient to allow any effects of that exposure to occur and be measured. The important difference between a cohort study and a RCT is that, in a cohort study, the groups are not randomised.

A good example of a cohort study is the Nurses Health Study, coordinated by Harvard University in Boston, in which almost 100,000 nurses agreed to fill out questionnaires mailed to them annually asking about their diet, health status, and so forth. Any reports of illnesses were then confirmed by medical reports. This study provided information about the factors that may be involved in many diseases, including cancer, diabetes, osteoporosis and heart disease. It was also the source of much of our earlier evidence that hormone replacement therapy (HRT) reduced your chance of heart attack. As we have mentioned a number of times in this book, when a randomised trial was finally done, it proved the opposite in older women!

**Case–control studies**

Case–control studies are observational studies in which a group of people who have a particular disease are observed to see whether their past exposures to some factors differ from those of a similar group who do not have the disease. They usually involve smaller numbers of people. They can be either population based – all the cases and controls are randomly selected from the same defined geographical population – or hospital based – all the cases and controls are selected from people attending a particular hospital. The population-based case–control study yields stronger evidence than one that is hospital based, which has many more sources of bias.

In case–control studies, information is usually obtained by questioning the cases and controls. Naturally they will know whether or not they have a certain disease, and this knowledge is likely to influence the way in which they respond to questions, which, in turn, can lead to recall bias – a type of measurement bias.

One important example of a population-based case–control study investigated whether the risk of sudden infant death syndrome (SIDS) is affected by a baby’s sleeping position and the quantity of bedding used. British researchers compared the situation of 72 infants who had died suddenly and unexpectedly (of whom 67 had
died from SIDS) with 144 control infants. They all were from a defined geographical area in the country – Avon and part of Somerset. The parents of the control infants were interviewed within 72 hours of the index infant’s death. Information on all babies was collected on bedding, sleeping position, heating and recent signs of illness.

Compared with the control infants, those who had died from SIDS were more likely to have been sleeping prone. As recall bias may occur in such a study (parents’ reporting may be influenced by what they have heard about the disease), it is important for such results to be confirmed by RCTs or, at least, by cohort studies. The results of cohort studies have since confirmed that the prone sleeping position increases the risk of SIDS. As it would not be ethical to randomise babies to different sleeping positions, a cohort study is likely to be the highest level of evidence available on this topic.

Cross-sectional analytical studies

Cross-sectional analytical studies are prone to more biases than cohort and case–control studies. These studies measure the two factors – exposure and disease – at the same time. For example, this method was used to see whether *herpes simplex* virus occurs more often in cervical cancer cells than in non-cancer cells. The problem with this approach is that it is unclear in which direction the causal arrow is pointing. We cannot tell from such a study whether the viruses cause cancer or whether the cancer cells are more susceptible to the growth of viruses.

Cross-sectional analytical studies are prone to selection bias. Consider what would happen if such a study were used to explore the relationship between cholesterol levels and the presence of coronary heart disease as measured by electrocardiography. There would be a problem in interpreting the results because, if high cholesterol does indeed cause a severe form of heart disease, by the time the study is ended and the sample population investigated, those with the highest levels of cholesterol are likely to have died from heart disease. In other words, this study will underestimate the relationship between high cholesterol levels and heart disease because it is based on a survivor population.
However, cross-sectional analytical studies are very useful for examining how well diagnostic tests identify the presence or absence of disease.

Opinions, case reports and anecdotes

Opinions, case reports and anecdotes all have one thing in common: they are largely based on personal experience that cannot be reliably generalised to other people. Their reliability as a source of evidence for an intervention is very weak except in highly specific situations that have to do with the nature of the illness being treated and the impact, immediacy and repeatability of the intervention’s effect (see Chapter 8 on anecdotal evidence).

An example: judging whether you should switch to a Mediterranean diet rather than a low-fat diet if you are at higher risk of heart disease

Observational (non-randomised studies) across a number of countries have noted that people who have a so-called Mediterranean diet have a lower rate of heart attacks and strokes and live longer. This seems plausible and, if you look for a randomised trial that tests this theory, you will find a recent one that randomly assigned people with higher than average risk of heart disease to either a low-fat diet or a Mediterranean diet. On first glance it looks like the Mediterranean diet was superior but there are some doubts raised when you look more closely.

First, the researchers have mainly used surrogate outcome measures such as lipid levels and blood pressure. The additional benefits of a Mediterranean diet over and above a low-fat one were statistically significant but possibly wouldn’t have much of an effect on clinical outcomes such as survival or heart attack rates. Also, they followed people for only 3 months so we don’t know whether there was any effect on heart attack rates or survival.

Second, when you look more closely, the Mediterranean diet group got more educational material and free supplies of virgin olive oil and nuts, whereas the low-fat diet group did not have any educa-
JUDGING WHICH TEST AND TREATMENTS REALLY WORK

tion and did not get low-fat products supplied free of charge. You should be able to see the potential for bias here. You might decide to switch to a Mediterranean diet because you prefer the flavours and because it is unlikely to do you any harm. However, there is no conclusive evidence that it will reduce your chance of having a heart attack.

If you wish to test your skills on appraising health information, using what you’ve learnt in this chapter, try examples 1–4 in Chapter 15.

Summary

When trying to assess the effects of tests and treatments, we should look for valid evidence that is relevant to our needs. This may be provided in an evidence-based guideline or systematic review of randomised controlled trials.

To ensure that evidence about a test or treatment is relevant:

- The evidence should describe outcomes that are person centred. They should tell you about survival or quality of life rather than surrogate measures.
- The evidence should describe both harms and benefits and tell you how likely they are.

Studies vary in their reliability as a source of evidence. The following is the hierarchy of evidence about interventions, from strongest to weakest:

1. systematic reviews
2. randomised controlled trials
3. non-randomised studies
   - cohort and other non-randomised trials
   - population-based case–control studies
   - hospital-based case–control studies
   - other types of studies
4. opinions, case reports and anecdotes.
CHAPTER 10

References

What makes you sick?

A universe in which cause and effect always have a one to one correspondence with each other would be easier to understand, but it is obviously not the kind we inhabit.

*Jerome Cornfield*

Living near power lines causes cancer. So does eating burnt steak. Stress at work will kill you. So might fluoridated water.

These are just some of the many recent claims about causes of disease. But how are these claims tested? We know that randomised controlled trials (RCTs) are best for testing the effect of interventions, but imagine randomly allocating people to live near powerlines so you could then measure the impact on cancer rates. Clearly, randomised controlled trials cannot be used in these situations.

So we have to look for other ways of assessing the influences of different exposures on health. ‘Exposure’ is the term used to describe a possible cause of disease such as something in our environment or a lifestyle behaviour. One way is to use randomised controlled trials to measure the impact of eliminating or reducing the exposure: measuring what happens to people’s health when they quit smoking compared with people who keep on smoking.

Sometimes randomisation occurs ‘naturally’. One study investigating the effects of wartime activity on combatants was naturally randomised because conscripts to the American army were selected...
by lottery on birth date. The findings showed an increased overall death rate, with much of the increase resulting from suicide and car accidents. The evidence from such a trial is stronger than from observational (non-randomised) studies.

When randomisation is impossible, observational studies can provide answers, based on observing what happens to people in various groups over a period of time while nature is allowed to take its course. Observational studies, such as cohort and case–control studies, differ from experimental studies in which some action is taken to change an outcome.

Observational studies are used:

- To investigate the causal link between an exposure or disease: for example, to follow a group of people to see whether those on a certain diet develop more heart disease or cancer than expected; or to observe mine workers over a time period to see whether they are at increased risk of lung disease.
- To follow up long-term rare effects: for example, a medication that has been approved on the grounds of an RCT showing...
short-term benefits can be followed up to assess its long-term outcomes by an observational study.

- To calculate risk predictors of disease: for example, to establish whether age, family history of breast cancer and the number of births are predictors of breast cancer risks.

**How to know when a relationship is causal?**

Even when there is an association between two factors – such as using a walking stick and grey hair, or eating carrots and car accidents – the one does not necessarily cause the other. To determine whether a condition or disease is actually caused by something we do or eat requires that several criteria are present.

<table>
<thead>
<tr>
<th>Does X cause Y?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The evidence about the relationship should be from a reliable source – the best possible study type.</td>
</tr>
<tr>
<td>2. The exposure to the supposed cause should occur before the outcome.</td>
</tr>
<tr>
<td>3. There should be a strong relationship between the supposed cause and the outcome.</td>
</tr>
<tr>
<td>4. There should be a dose–response or exposure–response relationship between the supposed cause and outcome, that is, the greater the exposure, the more likely someone is to get a disease.</td>
</tr>
<tr>
<td>5. There should not be any other factor that could explain the relationship.</td>
</tr>
<tr>
<td>6. The same results should be shown in several studies.</td>
</tr>
<tr>
<td>7. The relationship should make sense.</td>
</tr>
</tbody>
</table>

**Is evidence about the relationship from a reliable source?**

In other words, how good is the quality of the studies providing the evidence? The quality of evidence from various study types is given in Chapter 10. Obviously we can rely more on an RCT than on evidence from observational (non-randomised) studies. For example, an RCT showing that lowering blood pressure in people with high
blood pressure will reduce the risk of stroke is better evidence that high blood pressure causes strokes than a cohort study suggesting an association between high blood pressure and stroke. Cohort and population-based case–control studies would provide better evidence than hospital-based case–control studies.

**Did exposure to the supposed cause occur before the outcome?**

This is about the causal arrow. Obviously studies showing that the supposed cause is present before the outcome provide stronger evidence. For example, a cross-sectional analytical study showing that people who have had a heart attack are more likely to exercise does not constitute strong evidence that exercise causes heart attacks. It just means that there is an association between exercise and heart attacks. We can’t tell which one led to the other because the cross-sectional study design is like a ‘snapshot’ at a certain point and doesn’t tell you anything about the sequence of events over time. It is quite possible that people who had heart attacks started exercising because of it and that accounts for the association.

**Is there a strong relationship between the supposed cause and the outcome?**

If an observational study shows that smokers are nine times more likely to develop lung cancer than non-smokers, this is a strong relationship and is likely to be causal and not the result of biases in the study. But if a study shows that people with a certain exposure are 1.3 times more likely to develop a certain disease, this is not a strong finding and may well be explained by some bias in the study design and not reflect a causal relationship.

**Is there a dose–response or exposure–response relationship between the supposed cause and the outcome?**

The evidence for causality is stronger if it can be shown that, the higher the dose or the longer the time of exposure, the greater the
risk of the disease. For example, the risk of lung cancer increases progressively as the number of cigarettes smoked each day increases. In Chapter 12 we mention the dose–response relationship in the example about the relationship between fruit juice consumption and Alzheimer’s disease.

**Can the relationship be explained by anything else?**

Even in well-designed studies, there might be important differences at the outset between the groups being compared. Supposing a study is planned to find out if workers in a particular industry have a higher risk of developing lung cancer than the general population. There is already good evidence that smoking causes lung cancer so, if the workers in this particular industry happen to be heavier smokers than the population with which they are being compared, it would not be surprising if the study showed an increased incidence of lung cancer in that industry. The question is: can this increase in lung cancer be attributed to the working environment or is it the result of other factors? In this example smoking is a confounder, which may bias the results. But unlike many biases, confounders can be adjusted for in the analysis. Age is a very common confounder because disease rates are usually closely related to age. Consequently, differences in age between the ‘exposed’ and ‘non-exposed groups’ might well be the cause of any differences between them in disease occurrence.

Always check whether studies have identified and taken account of potential confounders that may influence the comparison. They can do this by showing that the potential confounders occur equally across all groups, or by using some statistical techniques to adjust for them. If the study does not state that potential confounding factors have been adjusted or controlled for, assume that it has not been done.

**Are the same results shown in several studies?**

The more studies that show the same sort of effect, the more confident you can be in the evidence. Just as there are systematic reviews
of RCTs on the effects of interventions, so too there are systematic reviews of observational studies on the causes of diseases. Similarly, we need to be sure that all the high-quality studies have been identified and used.

Does the relationship make sense?

The fact that some process seems to make sense is not, in itself, good evidence. It does, however, add weight if the criteria described above have been fulfilled. When an outcome is expected based on an understanding of biological mechanisms, this has more credibility than something that is contrary to biological expectations. We know, for example, that cigarette smoke reaches the lung, so information about a link between smoking and lung cancer is more believable than, say, suggesting a relationship between loud music and lung cancer. If you want to test your skill in assessing a causal relationship, turn to example 5 in Chapter 15.

Summary

When it is impossible to randomise people to study the cause or a predictor of disease, observational studies can provide useful information about the following:

- Whether exposure to some environment or behaviour causes a disease; for example, to see whether diet has an effect on heart disease or cancers; or whether certain occupations cause specific diseases.
- To follow up long-term rare effects; for example, to assess long-term outcomes of medication that has been approved on the grounds of an RCT showing its short-term benefits.
- For calculating risk predictors of disease; for example, to see whether the risk of breast cancer can be continued
WHAT MAKES YOU SICK?

predicted by factors such as age, family history of breast cancer and number of births.

Criteria to help you determine whether a relationship is causal include:

• The evidence comes from a reliable source.
• The exposure to the supposed cause occurred before the outcome.
• The relationship between supposed cause and outcome is strong.
• There is a dose–response or exposure–response relationship between the supposed cause and outcome.
• The relationship cannot be explained by any other factors.
• Other studies show the same results.
• The relationship makes sense.

Reference