III

Stories and studies
Earlier in this book, we discussed the smart health choice essentials, five questions for deciding how best to deal with a health problem. They are:

1. What will happen if I wait and watch?
2. What are my test or treatment options?
3. What are the benefits and harms of these options?
4. How do the benefits and harms weigh up for me?
5. Do I have enough information to make a choice?

This section deals with what to do if the answer to the last question is that you do not have enough information to make a choice. Sometimes this situation can arise if you have been given conflicting information, perhaps because different practitioners are citing different research results. A lot of poor quality research is used as a basis for what is sometimes euphemistically called evidence. Be warned that poor studies can provide evidence that is, at best, weak and, at worst, dangerously misleading.

In the shopping centre . . .

As Jenny guided her trolley along the aisle, she heard a familiar voice coming from the other side of the cereals. It was her cousin, Elise, chatting to someone.
... Celbequine, Jack. Wonderful for arthritis too. It’s really worked for my tennis elbow.... And it’s completely natural so it can’t do any harm. Just herbs and vitamins. It’s been scientifically proven to cure people.’

‘Is that so eh? Maybe I should try it. My leg’s really been playing up lately. So where do I get it? How much does it cost?’

‘To be honest, Jack, it’s not cheap. But I look at it this way:

‘Three months’ supply costs the same as a visit to my physio. Look, I’ve got a brochure here, tells you exactly how it works. It says: “Stimulates the body’s own immune system with a combination of herbs and vitamins and helps to relieve pain. Contains extracts of celery and barley, calcium, beta-carotene, vitamin A, vitamin D . . . .” Research shows that this blend of powerful ingredients prevents cartilage loss and slows down the progression of arthritis.” It also says: “May cause nausea, stomach pain or diarrhoea.”

Anyway, it works for me.’

Jack thanked Elise and promised to think about the tablets.

Jenny continued her shopping, thinking about what she had just overheard. It was something she often thought about: how easy it is to convince people that something works by describing how it is supposed to work. In her 8 years as a practitioner she had come across this phenomenon all too often. And not just with her patients – she had seen many pharmaceutical company representatives use similar arguments when trying to convince her to prescribe their products.

Now Jenny happened to know Jack, and she was worried that he might buy the tablets, which she knew he could not afford, having recently been laid off work. It occurred to her that the pain in his leg might not even be caused by arthritis; it could be sciatic pain arising from an old back injury. In addition, he had previously undergone surgery for a stomach ulcer, which meant that he may be at high risk of suffering the tablets’ side effect of stomach irritation. In short, Jenny knew that, even if these tablets did relieve Jack’s leg pain, the harms might outweigh the benefit.
Jenny bumped into Elise in the next aisle, and invited her for a coffee after they’d finished shopping.

How believable are the claims? . . .

Later, over coffee, Jenny admitted overhearing some of the conversation with Jack, and asked about the brochure’s claim that a trial reported 130 people were helped by the tablets. ‘But do they tell you how many people were treated altogether?’

‘Not that I recall,’ said Elise. ‘But why do you ask?’

‘Well, if 150 people were treated and 130 improved, that’s pretty good. But if thousands were treated and only 130 improved, it’s a different matter. That means a lot of people took tablets for no benefit, yet risking the known side effects, not to mention those that are not yet known. It’s important to know, not just how many got better, but how many were treated.’

Elise was intrigued: ‘I get what you’re saying, Jen, but the fact of the matter is that I felt better on those tablets, so they obviously work.’
'There are many reasons why you could be feeling better. You might just have improved anyway, which happens more often with health problems than many people realise. And people tend to feel better when they have taken some positive action, even if the action itself has no effect. The point is, when health claims are made about any product, the research supporting these claims must be valid and the potential benefits and harms should be clearly described so consumers can make an informed choice.'

**About the trial . . .**

After Jenny had seen her last patient later that day, she found Elise in her waiting room. Elise had rung the manufacturer to find out more about the trial, and was excited by the results. The 130 helped by the treatment were out of a group of 250, and the manufacturer said this meant about 50 per cent of people could expect to improve on the treatment. This is how the research was described to Elise:

During the first week of the opening of a new mall, free samples of Celbequine were handed out to shoppers who said they had joint pain or arthritis. Each sample of 25 tablets was enough for one week’s treatment, after which the shoppers were invited to re-order free supplies each month over a 6-month period. At the end of 5 months, there were still 525 people in the trial. They were asked to fill in a questionnaire to receive their final quota of free tablets. A competition offering a prize of gym membership was included with the questionnaire.

What transpired was that only 250 of the 525 questionnaires were completed and returned. Out of these, 130 indicated ‘considerable improvement’ within 6 months of starting the treatment. Elise calculated that 130 out of 250 meant about a 50 per cent follow-up rate, as stated by the manufacturer.

But Jenny was not so sure. ‘These figures could be important, but I’m still sceptical. First, you have to understand that pain from arthritis – or from many conditions for that matter – is not constant.
It often fluctuates. Many of my patients with arthritis have long periods relatively pain free. They just improve spontaneously without any treatment.’

Jenny scribbled some figures as she continued: ‘They told you 130 people got better out of 250 who filled out the questionnaires. What about the other 275 people? Maybe they couldn’t be bothered to fill out the forms because they thought the treatment useless? You can see what will happen if we include them in the calculations.’

‘Yes,’ said Elise. ‘It might mean that only 130 out of 525 got better. That’s … less than 25 per cent who felt any benefit. But, wait a minute, what if it worked the other way? What if the people who didn’t respond did feel some improvement but were then less likely to fill in the forms because they weren’t so conscious of their aches? That could mean that three-quarters of people felt some benefit. But how do we know which is correct?’

‘That’s the point exactly,’ said Jenny. ‘We don’t know. It’s very difficult to draw any firm conclusions from this sort of study. Suppose the prize of gym membership influenced who was likely to send in the form. What if it meant that people who were feeling fit and well were more likely to send it back? Or you could argue the other way: that the healthier, fitter people are less likely to be interested in such a prize because they already go to the gym. Either way, you could argue that the study was more likely to include certain types of people. This is what researchers call selection bias, and makes it difficult to know whether the treatment really works.’

Elise interrupted: ‘So what you’re really saying is that we don’t know whether the tablets made 25 per cent of people feel better or 75 per cent of people feel better?’

‘And we don’t even know that,’ continued Jenny. ‘Remember that many hundreds more people, maybe even thousands, were given samples. Why did all those other people drop out of the survey? Maybe they suffered stomach pains or other side effects. Again, this is selection bias. As we don’t know what happened to everyone in the trial, we have no way of knowing whether the tablets worked.’
Some things get better on their own . . .

Jenny could see that much of this was new to Elise. She wanted to finish her train of thought, so she went on: ‘There is a very important phenomenon that is often overlooked. Researchers call it spontaneous remission, but all it means is that, in many instances, time heals. Our bodies have a marvellous capacity for recovery. For a lot of conditions, people get better without any treatment.

‘So getting back to the question of whether there is good evidence that Celbequine is effective: how do we know if the people who reported improvement were responding to these tablets or perhaps to something else that they were taking, or whether the pain just got better on its own, which might have happened even without Celbequine? Don’t forget that these are people who were well enough to be walking around a shopping centre when the trial began, which means their symptoms may not have been too bad to start off with. If that was the case, there’s a good chance they may have recovered on their own, with no treatment at all.’

The placebo effect . . .

Reaching for a book from the shelf, Jenny continued: ‘Now there’s something else that often gets in the way when you’re trying to judge the effects of a treatment. Let me read you something from this book by Norman Cousins called Anatomy of an Illness as Perceived by the Patient. It’s a marvellous account of how he dealt with a very serious disease in a most unconventional way.

‘This is the part where he’s describing a placebo:

“A striking example of the doctor’s role in making a placebo work can be seen in an experiment in which patients with bleeding ulcers were divided into two groups. Members of the first group were informed by the doctor that a new drug had just been developed that would undoubtedly produce relief. The second group was told by the nurses that a new experimental drug would be administered, but that very little was known about its effects. Seventy per cent of the people in the first group received sufficient relief from their
ulcers. Only 25 per cent of the patients in the second group experienced similar benefit. Both groups had been given the identical ‘drug’ – a placebo."

Elise thought for a moment. ‘So the implication is that the people in the trial could have been improving just because they were told they would – because they believed the tablets worked.’

‘Yes, there’s certainly the strong possibility that some were responding to the placebo effect,’ said Jenny. ‘The mind has mysterious powers. Sometimes believing is seeing! The placebo effect is very helpful but we want to know whether an intervention that has some risks and costs has an effect over and above its placebo effect.’

And other study flaws . . .

‘And there’s another problem with the way that this study was done. Having accepted free samples of the product, I bet not many people could have said it did absolutely nothing to make them feel better, let alone that it made them feel sick. I’m not saying people deliberately lie, but there’s often a temptation to be more positive than one might genuinely be feeling. This is sometimes called “acquiescence bias” and is an example of measurement bias. Imagine you’re in the supermarket and you’re offered a sample slice of a lemon meringue pie that you accept. The salesperson asks how you like it, with one of those smiles that says “isn’t it just too delicious for words”. Many people would find it difficult to say otherwise.

‘So getting back to the questionnaires, we’ve seen there could be several biases: one in the way people interpret and report their health outcomes, called measurement bias, and another in the exclusion of who knows how many hundreds of people who originally entered the trial – selection bias. The bottom line is that there is still no sound evidence that Celbequine does more good than harm.’

Elise was a little hesitant: ‘I follow what you’ve said, Jen, but are you telling me there is absolutely no value in the testimonies of those 130 people who thought the treatment helped them?’
‘What I’m saying is that when you’re making an important decision about whether some treatment is effective or not, your judgement should be based on stronger evidence than the personal testimonies of just a few people who took the intervention. This is true whether you’re a researcher doing a study on the effect of an intervention, or a consumer, or a practitioner advising a patient.’

Personal experiences can be important . . .

‘But your question is valid. Are anecdotes based on personal experience ever valuable? The answer is most certainly yes. If you experience a dramatic, immediate change in your symptoms after some treatment for a condition that usually lasts a long time without any treatment, and you experience the same, strong, rapid effect on several subsequent occasions, then your experience provides evidence of the treatment working for you. Let’s say, for example, you have regular migraine attacks that usually last several hours and a new tablet stops the pain within half an hour every time you take it. There’s little doubt the tablet is working, for you anyway. But for most medications, the effects are not that dramatic. For most medications, the improvements we are looking for are more subtle and occur over a longer period. This is where individuals’ reports are of little value.

‘Getting back to Celbequine; the information said it would take up to 6 months to work and also said something about retarding the development of arthritis long term. So in this case, individual reports are not a reliable guide to its efficacy. Remember that individual reports of improvement are no more than that – reports of how people feel after treatment. They do not tell us anything about what may have caused the improvement – whether it was the treatment, something else happening in the person’s life or just the passage of time.

‘Individual reports are of little use when we want to know what some medium-term or long-term change is caused by. That’s another story entirely. To make this kind of deduction, we need probabilistic data, that is, information about the percentage of people who improved. Moreover, we need to compare this percentage with the percentage who would have improved without the treatment.’
Let's get sceptical . . .

Jenny continued slowly and deliberately: ‘So the marketers of Celbequine might have done themselves a favour – if their product truly is as good as they claim – by doing their homework before embarking on a costly bit of research that was clearly full of weaknesses. On the other hand, if their claims are unwarranted, they might not want people to know. At the end of the day, if a product’s claims are genuine, well-designed research can only strengthen the claims, whereas, if there is no valid evidence to back their claims, we should remain sceptical.

‘To make well-informed choices about important decisions that may affect our health – our bodies and our minds – we need much more than opinions. We need evidence.’

They were both silent for a moment, then Jenny said: ‘But if you want my opinion, I think it’s time to call it a day!’

From detergents to treatment for acne . . .

When they met again a few days later, Elise was excited as she described her investigations of a new detergent that had been advertised as a breakthrough in ‘enzymatic power’ that ‘gets whites whiter than white’.

As it was more expensive than her regular powder, Elise wanted to test the claim. She divided her whites into two, and used her regular detergent for one load and the new product on the other.

‘And then I thought that this is the way that Celbequine should have been tested,’ she said. ‘Get two groups of people with the same health problem; give one group the treatment and give the others nothing or the old treatment.’

‘Bravo,’ said Jenny. ‘You’ve hit the nail on the head. But to take it a step further, imagine that you were setting up a study for a new acne treatment. What would you do?’

‘You need a bunch of people with acne,’ replied Elise. ‘Teenagers. You could approach high schools or advertise in teenage magazines asking for volunteers. And then divide them into two groups. One group is given the treatment and the other . . . what do
they get? I mean what’s in it for the volunteers if they’re not getting the new treatment?

‘Good point,’ said Jenny. ‘What say we tell the teenagers that we’re testing a new acne lotion. We don’t know whether it works or not but you could help us find out by taking part in a trial. The trial will work like this: we spin a coin – heads, you get the treatment, tails you get some other lotion that looks, feels, smells like the treatment but is inert, a placebo. If the new treatment is shown to work, we will offer a free course of treatment at the end of the trial to everyone who was given the placebo.

‘Doing it this way, you see, excludes other variables that might affect the outcome. For instance, the massaging action of applying the cream might do some good – or some bad for that matter.’

Elise was hooked. ‘And remember what you read me about the placebo effect? If the volunteers aren’t told whether they’re getting the treatment or placebo, this will stop their expectations influencing the results. No measurement bias, right?’

**Randomised and blinded . . .**

‘Right! And there’s another thing. In a comparative study, it’s essential that the groups are similar if you want the results to be valid. A good way to ensure this is to allocate the treatment randomly. You’ve heard of randomised controlled trials, haven’t you? Well, randomisation addresses the possibility that those who did not get the treatment were sicker than those who did – or vice versa. Randomisation, or random allocation as it is sometimes called, can be done by the toss of a coin, or by other techniques – computers can be used to allocate patients randomly to a treatment or placebo.

‘Each volunteer is randomly allocated to the new treatment or to the placebo; in addition, as you’ve already said, people shouldn’t know what group they’re in. In other words, they should be masked or “blinded” to whether they are getting the treatment or placebo. This is the way drug licensing authorities assess new claims about the effects of drugs.’

The two women sat in thought for a while. Then Jenny went on:
Listen Elise, obviously you and I can’t rush off and do a randomised controlled trial on every new pill or powder on the market, but the healthcare system is continuously involved in studies of all types. There are researchers out there doing randomised controlled trials all the time … or RCTs as they are sometimes called. What consumers should be doing, however, is asking their practitioners for evidence supporting their recommendations about treatments or any tests – especially for important decisions. The onus is on the person or organisation that recommends the product or the treatment or the service to supply sound evidence that it improves or prolongs life. This includes practitioners from all areas and doctrines of healthcare, and the pharmaceutical companies who recommend their products.

‘Thousands of studies are published every year – though not all of them are randomised controlled trials – and are accessible to practitioners either in journals or in summarised form on electronic databases. So when evidence is available, it should be used.’

Do the benefits outweigh the harms? ...

‘If I were approached by representatives of this company to recommend Celbequine to my patients, I would expect them to provide me with sound evidence that my patients are going to be better off. Failing that, I could do a computerised search to see whether there is any evidence in the medical literature. I could find Medline on the internet, for example, and look at the abstracts of studies published in the most important journals in recent years. Or I could look up the Cochrane Library, which is a regularly updated electronic library of summaries of all the randomised controlled trials (see page 138). The point is that, if I have no valid evidence that the benefits of a treatment outweigh its harms, I should be careful about recommending it. Most practitioners now have access to online computer systems in their offices to search for valid evidence. Mind you, being available on the internet means it can be accessed by anyone.’

Elise thought for a bit. ‘Why didn’t the manufacturer do the study right in the first place?’
'Now that,' said Jenny, 'is a very interesting question. Maybe they don’t know about randomised controlled trials or maybe they think consumers won’t know the difference. It’s just possible, of course, that the stuff doesn’t work and they prefer not to make that knowledge public. Or maybe, good evidence is simply not available. After all, randomised trials can be complex and expensive to conduct.

So does it really work? . . .

‘But let’s get back to the question at hand: does Celbequine really work? The only information that we have is not very convincing because it is based on individual reports. This can be misleading for many reasons: first, people often improve spontaneously with time; second, they might be responding to the placebo effect. Of course, it could be that the treatment really does work. But we can’t judge this from the information available. And yet we have this situation where expensive, potentially dangerous interventions are recommended without valid evidence that they work. What we need is good research to supply valid evidence so that all of us can make informed judgements about our health.’

Elise thanked her cousin and left with a mixture of new-found confidence as well as surprise that she had been so poorly informed before. Why, only a few days ago she would have thought herself as well informed as the next person. Then she realised, with some surprise, that she probably had been.

Note: the name ‘Celbequine’ is purely fictitious.

Reference

The strength of one

Theists, for example, note the number of times their prayers have been answered and conclude that there is a benevolent god; atheists cite the occasions that their prayers have gone unanswered and conclude that we are on our own. Both need to develop the habit of thinking more broadly. Both must consider the number of times their hopes have been answered when they have prayed and when they have not, as well as the number of times their hopes have been dashed when they have prayed and when they have not.

*Thomas Gilovich*

One of the points raised in Chapter 7 is the compelling allure of personal testimony. For many of us, this is one of the most seductive sources of health information. A neighbour says her cancer disappeared after she took shark cartilage. Your mother swears that taking a vitamin C tablet every morning keeps her free of colds. A colleague claims his bad back recovered after doing a certain exercise for 6 weeks.

It can be tempting to draw conclusions from such anecdotes; somehow a story involving a real person whom you know can seem more convincing than the results of studies based on thousands of anonymous participants. Anecdotal evidence is usually based on individual experiences or observations, as distinct from probabilistic evidence that gives estimates of how likely something is to occur.
based on experience with large numbers of people. In this chapter we discuss some of the ways in which stories can be helpful in making health decisions, but also warn of their limitations.

**The danger of the anecdote**

There are inherent dangers in relying totally on anecdotes. Consider the case of Mr Dickens, 70, who recently consulted Dr Carter about an irregular pulse. Mr Dickens, who has previously had high blood pressure and a stroke, is found to have a disturbance of his heart rhythm, called atrial fibrillation. This condition may cause a blood clot to develop in the heart and send off fragments that can cause a stroke by blocking arteries in the brain. One treatment used in people with atrial fibrillation is an anticoagulant, which prevents a blood clot. But Dr Carter knows that this drug can also cause internal bleeding, with potentially serious consequences, although she has never had a patient suffer this particular side effect. After she prescribes the treatment, Mr Dickens has a bleed into the brain.

Soon after, Dr Carter sees Mr Jones, another elderly man with similar problems to Mr Dickens. But Dr Carter does not prescribe an anticoagulant this time, discouraged by her recent experience with Mr Dickens. Mr Jones later suffers a stroke. Dr Carter will never know if this might have been prevented if she had prescribed an anticoagulant. But an examination of the probabilistic evidence – as distinct from the anecdotal evidence provided by case reports – gives us some idea.

If Dr Carter had done a literature search, she would have found several good randomised controlled trials showing a two-thirds reduction in stroke for patients treated with anticoagulants. On the other hand, serious bleeds from anticoagulation are rare, so overall her patients with atrial fibrillation would be served best by taking anticoagulants unless they are at low risk of stroke or at high risk of bleeding. Mr Dickens is at high risk of stroke because he has high blood pressure and has had a previous stroke. Out of 1000 people like Mr Dickens who are treated, about 120 strokes would be prevented in the next year, whereas a bleed into the brain as a result of anticoagulants would occur in about 5 people. In addition, there may be other bleeding, some of which would be mainly a
nuisance such as bruises, and some of which could be more serious, such as bleeds into the stomach or bowel.

It [an anecdote] is useful for documenting that the outcome can occur, but provides no information about the frequency with which it occurs or the effect of an intervention on the frequency of occurrence.

*David Eddy*

**Only survivors speak!**

Anecdotes have limited use in judging the effectiveness of health interventions. If you wanted to know, for example, whether a certain cancer treatment saves lives, the opinion of someone who had the
treatment would not be a reliable guide. Remember that those patients in whom the treatment did not work are no longer around to give their views. Only survivors speak – which can result in a very biased picture of an intervention.

**More problems with anecdotes**

Another problem with anecdotal experience is that we tend to give the most recent and negative experience undue bias. As a result of this phenomenon, most of us are inclined to be over-confident when making predictions based on a recent experience, even when we have more reliable probabilistic information on hand. It is therefore especially inadvisable to use anecdotal evidence to assess a treatment with long-term effects.

Here are the reasons why anecdotal evidence is weak when judging most interventions:

- The outcomes of most health problems are not predictable for any individual. How a health problem will affect an individual is difficult to predict and can be expressed only as a probability. For example, you may have a 40 per cent chance of surviving for another 10 years. An intervention can be judged only by the extent to which it changes this probability of survival. Just because you are alive at the end of 10 years does not mean that the intervention is responsible.
- The effects of most interventions are small and subtle. An intervention may increase the chance of living for 10 years from 40 per cent to 50 per cent. It would be impossible to detect such an improvement based on anecdotal reports.
- The effects of many interventions are long term. It is difficult to link an outcome – whether that happens to be a recurrence of disease, good health or death – to an intervention used years before. There may be a host of other factors involved.
- The effects of some interventions cannot be confirmed by testing the intervention on yourself again. If you suffer from migraine, which usually causes a persistent headache, and this symptom disappears as soon as you take a certain tablet, you
THE WEAKNESS OF ONE

can test this hypothesis next time that you get a migraine. But most conditions do not recur repeatedly, so you have no opportunity for confirming the effect of an intervention.

But anecdotes can be useful in some situations

It is generally unwise to rely on other people’s stories as a guide to how likely you are to experience similar benefits or harms from an intervention. However, anecdotes are useful in some situations.

When confronted by illness or other health problems, many people find it helpful to talk to others who have been through similar situations. Their stories can provide useful insights into how your life might be affected by a similar illness, or side effects from a treatment, and what strategies might be useful in helping you deal with them. Indeed, some universities now invite patients to talk to medical students about their own experiences with illness, in an attempt to ensure that doctors become more understanding of and sympathetic to what it is like to be a patient. Other people’s stories can also provide useful information on how to find your way around the health system, which can seem like a confusing maze to a newcomer. Patient support groups can be particularly useful in these situations.

Most scientific and medical discoveries have their roots in anecdotes, which have led to hypotheses that are then proved by rigorous testing. In some circumstances, the anecdotal evidence can be so spectacularly convincing that the need for further confirmation diminishes. For example, when Howard Florey and Ernst Chain developed the drug penicillin based on Alexander Fleming’s earlier work, the antibiotic properties were so striking that it was introduced for use without long-term trials. When people are treated for an illness and survive in the face of evidence that most people die without treatment, there is usually little doubt about the treatment’s efficacy.

Let’s explore a situation where personal experience may help to decide whether a treatment works. Consider the case of Ed Smith, who suffers from severe migraines. When a migraine strikes, Ed is incapacitated and has to lie down for several hours. He has tried
many supposed remedies over the years, but none has worked. He
hears from a friend about a new therapy that has helped her. He tries
it the next time he feels a migraine starting, and his pain disappears
quickly.

Is Ed’s excitement about the treatment well founded? He
knows that the pain disappeared quickly when he took the tablets,
and that they helped his friend. He now also needs to know whether
he will experience the same immediate benefit when he takes the
tablets again. If indeed he tries it again and he experiences the same
pain relief, he probably has good reason to feel excited.

As with the introduction of penicillin, anecdotal evidence can
be used to assess the affect of a treatment if at least some of several
principles are fulfilled:

• The outcome of the disease or condition is predictable in the
  absence of the treatment. The condition in question does not
  usually get better on its own, at least not immediately.
• The effect of the treatment is immediate. The outcome is evident
  soon after the treatment.
• The effect of the treatment is large. There is a dramatic, large
  and obvious effect that would be difficult to attribute to sponta-
  neous improvement.
• The effect of the treatment can be confirmed by repetition. If
  the nature of the condition is such that it recurs, it is possible
  to confirm the treatment’s effects by repeated testing.

Acting on someone else’s anecdotal experience is appropriate only
if the harm seems small and the benefit worthwhile. Suppose that
you suffer from chapped, itchy skin in winter, and a friend tells you
about a new cream that helped him. Should you try it? It sounds
like you should. First, if it relieves the dryness and itch, you can be
fairly sure the cream is responsible if previous treatments have failed
to make any difference. Second, because treatment is likely to be
short term, the risks of serious adverse effects are low. Third,
judging whether the cream is effective is straightforward. So, if the
cream works, you will benefit and, if it doesn’t, you stand to lose
very little.
N of 1 trials

We could be more scientific about assessing the effect of the anti-itch cream, even with just one person, by using what we call an ‘N of 1’ trial. These trials have been defined as:

… [a trial where] the patient undergoes pairs of treatment periods organised so that one period involves the use of experimental treatment and the other involves the use of an alternate or placebo therapy. The patient and physician are blinded, if possible, and outcomes are monitored. Treatment periods are replicated until the clinician and patient are convinced that the treatments are definitely different or definitely not different.

If we had a ‘fake’ or placebo cream as well as the active one and tried each, one at a time, without realising which one we had tried, this would be an ‘N of 1’ trial. Ed Smith could have done one of these for his migraine treatment if he did not have labels on the pills and tried the new one against a fake one without knowing which was which.

In short, anecdotal information is useful when you are looking for immediate symptomatic relief for a relatively minor condition, and there is little potential for the treatment to do harm. It is also useful if you want to know how other people coped with a specific problem, or gain some insight into their experiences of diseases or interventions. It might be helpful for generating hypotheses that can be more rigorously tested.

A broad range of stories

But what if we have a collection of stories on the same topic? This may help us to learn about the context of what can happen in an illness or the likely sequence of events. For example, the same strain of flu can render one person sick in bed for a fortnight whereas another manages to keep functioning and is better within 4 or 5 days. People’s stories can be used to develop concepts and hypotheses, which can then be more rigorously tested and evaluated with a large group of people and in different settings.
The usefulness of a broad range of stories depends on how they have been collected and analysed. On most topics, if you listen to enough people’s stories, eventually you will start to hear similar accounts of the illness or treatment in question. As a very rough guide, this usually happens when you start to collect more than about 30 stories and you can be reasonably confident that you’ve captured the most common and likely experience of an illness or its treatment. Ideally two different researchers should look at recordings of the stories and identify the common themes within them.

A good example of this is a website called DiPEx (Database of Individual Patients’ Experience) at www.dipex.org. This website has over 100 modules on different illnesses and patient experiences. Each module consists of a number of patients’ stories that typify over 40 or 50 stories that were recorded on each topic. This means that the main patient experiences are more likely to be covered.

A resource like this is more powerful and useful than just one anecdote when you are trying to make a decision, because it is a bit more balanced and provides a range of experiences from a number of people, not just one perspective or opinion. However, although it gives a range of experiences, it does not provide information on how commonly they occur.

**Summary**

As seductive as anecdotal reports can be, it is usually unwise to rely on generalisations based on one or two experiences. They do not tell us the most probable outcome, which is most useful for guiding decisions. Anecdotal evidence is useful to help you understand the nature of the symptoms of a disease and of the side effects of treatment. However, anecdotes are poor evidence of how likely that outcome is to occur, except in a few circumstances as shown below:

*continued*
The weakness of one

Table 8.1 Comparison between reliable and unreliable use of anecdotal evidence

<table>
<thead>
<tr>
<th>Anecdotal evidence is reliable</th>
<th>Anecdotal evidence is unreliable</th>
</tr>
</thead>
<tbody>
<tr>
<td>When the outcomes of the disease or condition are predictable in the absence of treatment (e.g. migraines, chronic arthritic pain, premenstrual tension)</td>
<td>When the outcomes of the treatment are uncertain for the individual (e.g. breast cancer, diabetes)</td>
</tr>
<tr>
<td>When the effects of the treatment are large</td>
<td>When the effects of the treatment are small and subtle</td>
</tr>
<tr>
<td>When the effects of the treatment are immediate</td>
<td>When the effects of the treatment are delayed</td>
</tr>
<tr>
<td>When the effects of treatment can be confirmed by repetition</td>
<td>When the effects of the treatment cannot be confirmed by repetition</td>
</tr>
<tr>
<td>When the effects of treatment can be confirmed by an ‘N of 1’ trial</td>
<td>When the effects of treatment are disproved by an ‘N of 1’ trial</td>
</tr>
</tbody>
</table>

References

… while the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician.

Sir Arthur Conan Doyle

The power of probabilities

We live in an uncertain world where, as is so often said, the only certainties are death and taxes. Even so, learning to ask the right questions can help reduce the uncertainties that surround health outcomes.

One of the important questions relates to the probability of events, such as the chances that an intervention will cause a particular benefit or harm. Although anecdotal evidence is generally unreliable because it is based on an individual’s experience, probabilistic evidence is more reliable because it is based on the experiences of many and therefore tells you what is likely to happen. Probabilistic information is derived from studies involving many people. When assessing the probable outcome of a test or treatment, the most reliable probabilistic information comes from randomised controlled trials (RCTs) or systematic reviews of RCTs.
Probabilities – what do they tell us?

When making a decision about how to manage any health problem, probabilistic evidence answers the following questions:

1. How likely is a particular outcome?
2. What factors affect the chance of this happening?
3. How will a particular test or treatment change the chance of this happening?

How likely is a particular outcome?

To estimate how probable an event is, numbers are far better than verbal assurances or clichés. Suppose you are trying to assess the
chances that a healthy 5 year old will suffer complications after a tonsillectomy. If your practitioner’s response is ‘I’ve never lost a patient from tonsillectomy’ or ‘the risks are small’, you need to ask for probabilistic information. After all, what your practitioner considers a ‘small risk’ may be an unacceptable risk to you.

Far better to avoid misunderstandings through the use of data such as: ‘one in 100 people has severe pain after this procedure’ or ‘one in 1000 people will need further surgery’ or ‘one in 100,000 people dies from this procedure’. If you are not given probabilistic data, ask for it. And if you do not understand what the figures mean, ask for them to be explained in terms that you do understand.

What factors affect the chance of this happening?

Many women are worried about their risk of developing breast cancer. Suppose that you hear on the news that the latest figures show that 91 in every 1000 women are at risk of breast cancer (that is, 1 in 11).² This sounds alarming. However, remember that a woman’s age has a powerful effect on her risk of developing breast cancer at a particular time.

It has been estimated, for example, that 13 in every 1000 Australian women will develop breast cancer between the ages of 40 and 50, compared with 24 in every 1000 women between the ages of 60 and 70.³ For breast cancer, another important risk factor – so named because it increases a woman’s chance of developing the disease – is having a strong family history of the disease. Although women without such factors are at lesser risk of developing the disease, this does not mean that they are risk free. And, conversely, women with many risk factors for the disease will not necessarily develop it.

Being older affects people’s risk in many situations. For example, elderly people are more likely to suffer serious complications from influenza than younger people. This is one of the reasons why annual flu immunisation is recommended for elderly people.

When making health decisions, you need to know what factors affect the probability of you having a particular outcome – whether it changes your chance of suffering nasty effects from a disease, test
or treatment. Such factors could include your age, and your medical and family history.

**Will a particular test or treatment change the chance of this happening to me?**

The only way to find out whether a test or treatment will change the likelihood of an outcome is to turn to the results of studies that have compared the outcomes of people who received the test or treatment with the outcomes of those who did not.

It is this sort of study that tells us that children who have a persistent green runny nose will be more likely to recover if they have 10 days of antibiotics than those who do not. A Cochrane systematic review of six randomised controlled trials involving a total of 401 children recorded the number of children who were cured after randomly allocating some to receive antibiotics and the others to receive placebo pills. No-one knew whether they were in the intervention or placebo group.

After finishing their course of tablets, 56 in 100 of the placebo group still had runny noses, compared with 40 in 100 of the antibiotic group. However, 2 in 100 children who were taking the placebo tablets reported side effects compared with a slightly higher 5 in 100 on the antibiotics.

Clearly, this study does not show that giving children antibiotics for persistently runny noses guarantees that they will be cured, but it increases the probability that they will. As we learnt earlier in this book, you might also have noticed that a fair number of children will get better without antibiotics too.

To find out more about how the effects of treatment are represented numerically, see Chapter 18 describing relative risk and risk difference.

**Putting probabilities to work**

**Example 1: Anticoagulants for atrial fibrillation**

Let’s return to the example of Dr Carter from Chapter 8. Dr Carter’s patient was an elderly man with atrial fibrillation (a disturbance of
heart rhythm), and she was considering treating him with anticoagulants, which prevent blood clots forming. But Dr Carter also knew that anticoagulants can cause internal bleeding, so she had a dilemma. What a practitioner in her situation should do is to find out the probability of preventing a stroke by treating with anticoagulants, as well as the probability of serious side effects from anticoagulants.

If she had searched the medical literature, she would have found the following probabilistic information produced by the American College of Chest Physicians.

What is the probability that an elderly person with atrial fibrillation will have a stroke?

One option is not to take any medication for atrial fibrillation. Without medication, an average patient with atrial fibrillation (not caused by rheumatic fever) has about a 5 per cent chance of suffering a stroke in the following year (10 per cent over 2 years).5

What factors affect the probability of an elderly person with atrial fibrillation having a stroke?

As for most diseases, there are certain characteristics or risk factors that help to determine who is more likely and who is less likely to have a stroke in this situation. In this case, the risk of stroke increases with age, high blood pressure, some heart failure and any previous history of ‘funny turns’ or strokes. For example, if you are aged under 65 years and have none of the above risk factors, the chance of you having a stroke over the next 2 years is 2 in 100. This can be shown diagrammatically by the 100 faces in Figure 9.1: 98 of them do not have a stroke, one has a mild stroke (light shaded) and one has a severe stroke (darker shaded).

On the other hand, someone who is over the age of 75 years and has atrial fibrillation, and who also has high blood pressure, has a much higher risk of stroke over the next 2 years – 20 out of 100. Figure 9.2 shows this graphically.

Will anticoagulants change the probability of a stroke?

The answer is yes. Randomised controlled trials have shown that thinning the blood with either aspirin or the stronger drug
Warfarin can reduce the chance of stroke in people with atrial fibrillation. The risk of stroke is reduced by about 65 per cent with warfarin and by about 22 per cent with aspirin. This is because blood clots can form in the fibrillating heart chamber, break off and go up to the brain, causing a stroke. The higher your risk from the outset, the greater your benefit from treatment. A 65 per cent reduction if your risk starts out at 20 in 100 is going to mean it comes down to about 7 in 100, whereas if your risk of stroke starts out at 2 in 100 then warfarin will reduce it by 65 per cent to around 1 in 100.

**Figure 9.1** From *Making choices: treatments to prevent strokes in patients with atrial fibrillation.* (Light grey = minor stroke, dark grey = major stroke)
However, it’s not all good news because thinning the blood in this way also increases your chance of bleeding, which can be serious if it comes from a stomach ulcer or your brain, for instance. The chance of serious bleeding when taking warfarin is about 4 in 100 people. So, it’s a trade-off between lowering your risk of stroke (bearing in mind how great your risk is from the outset if you do nothing) and your risk of stroke. Warfarin lowers your stroke risk by a greater amount but is more likely to cause serious bleeding. It seems that, in this case, as in most of life, you don’t get something for nothing!

Any benefit of anticoagulants is ‘bought’ at a cost of an increased chance of a life-threatening bleed such as a stomach bleed (shown as black shading in Figure 9.3), in addition to minor bleeding such as
Figure 9.3 From Making choices: treatments to prevent strokes in patients with atrial fibrillation. Version for people with low stroke risk. (Light grey = minor stroke, dark grey = major stroke, black = serious bleeding)
brui

Figure 9.4 From Making choices: treatments to prevent strokes in patients with atrial fibrillation. Version for people with high stroke risk. (light grey = minor stroke, dark grey = major stroke, black = serious bleeding)
no risk factors for stroke, in whom taking anticoagulants will reduce his or her chance of stroke from 2 per cent to 1 per cent.

But for someone with all three risk factors for stroke, the price is likely to be outweighed by the benefit of reducing the chance of a stroke from 20 per cent to 7 per cent (Figure 9.4).

Aided by our practitioners, we need to exercise judgement in deciding how to apply probabilistic evidence to our own decision-making. For example, if you have atrial fibrillation and are at high risk of a stroke, anticoagulants are probably a good idea because the benefits outweigh the potential harms. But if you are at low risk of stroke, and especially if you have a high risk of bleeding, anticoagulants may be best avoided because the harms outweigh the benefits. If your risk is moderate, and the risks and benefits are less clear-cut, individual preferences become particularly important. The critical factor to consider here is how the individual weighs the ‘value’ of preventing a stroke against risking a bleed and other disadvantages of taking anticoagulants, including the tendency to bleed and the need for regular monitoring.

The figures used to illustrate these probabilities come from a Canadian decision aid that is based on high-quality randomised controlled trials. It has four different versions, depending on your risk and can be found at www.canadianstrokenetwork/research/clinicians.php

Example 2: Hormone therapy after the menopause

Another common example where the probabilistic evidence needs to be carefully weighed is the use of hormone replacement therapy (HRT) after the menopause. In this case, women need to weigh up the short-term benefit of symptom relief from hot flushes and night sweats, against the longer-term risks of breast cancer, abnormal mammograms, blood clots and strokes. There also appears to be an increase in risk of heart attack during the first year of taking HRT.

As we mentioned earlier in this book, the results of a large well-conducted randomised controlled trial, the Women’s Health Initiative (WHI), were published in 2003 and overturned some of our previous beliefs about HRT.
The chance of still having hot flushes and night sweats 12 months after starting HRT are 233 in 1000 women compared with 482 women in 1000 who took a placebo for 12 months. In other words, almost half of women aged 50 will naturally have fewer symptoms after 12 months but almost three-quarters will get relief if they take HRT. Figure 9.5 is based on extracts from a decision aid based on the evidence from the WHI study commissioned by the National Health and Medical Research Council (NHMRC), Australia.8

On the other hand, this same study showed that over 5 years a woman’s risk of developing breast cancer increased from 11 in 1000 to 15 in 1000 on HRT, the chance of having an abnormal mammogram increased from 84 in 1000 to 139 in 1000, the risk of stroke increased from 4 in 1000 to 6 in 1000 and the risk of serious blood clots increased from 3 in 1000 to 8 in 1000.

Your probabilities need to be weighed up, together with what is most important to you. Decision aids for consumers, such as that

---

**Figure 9.5** From *Making decisions: should I use hormone replacement therapy?*8 [2 gold star evidence means a high quality randomised controlled trial]
from which the Figures have been reproduced, have been shown to help people become more actively involved in their healthcare decisions. An example of a balance sheet or summary of these probabilities along with what we don’t know about HRT is shown in Figure 9.6.

For other examples on the internet, see the patient decision aids designed by the Sydney Health Decision Group at www.health.usyd.edu.au/shdg and the Ottawa Health Research Institute http://decisionaid.ohri.ca/decaids.html.

Using probabilities to balance benefits and harms

The more we understand the probability that a test or treatment will cause a particular benefit or harm, the more informed our decision-making can be, as shown in Figure 9.7, and the more certain we can be of the results of our decisions.
Balancing benefits and harms of tests and treatments

Probabilistic thinking helps protect us from the barrage of diagnostic tests and treatments advised by the media, friends and family. It also helps us to evaluate better the advice of our practitioners, whose concern about litigation may lead them to recommend unnecessary tests. Indeed, if we and our practitioners had more open discussions about the probable outcomes of different interventions, this might help reduce litigation.

When people are led to expect a definite answer, a definite cure, they may quite understandably blame each other when things go wrong. The malpractice suit is the patient’s way of blaming the doctor; the charge of ‘non-compliance’ is the doctor’s way of blaming the patient. Under the Probabilistic Paradigm the fact that things may go wrong, and that it may or may not be anybody’s fault, is acknowledged from the start.

Burstajn et al.¹⁰

Practitioners should be judged on the process of care, rather than necessarily on the outcome of a disease or treatment. If a practitioner does the reasonable thing by declaring the risks and their probabilities based on the best available evidence and thereafter acts on this information, taking account of patient preferences, he or she has done the reasonable thing, regardless of the outcome.
THE POWER OF MANY

Having information about probabilities allows us to decide if the dangers of a test or treatment outweigh their potential benefits. Ask your practitioner about probabilities; question how likely a good or bad health outcome might be. Think probabilistically when you read or hear health advice. By learning how to use probabilistic information, you will have an idea of what outcomes you can expect if you are ill and whether any interventions are likely to make a difference.

Summary

To estimate your chance of recovering from a disease, or of being helped or harmed by a test or treatment, you need probabilistic information. Taking a probabilistic approach to health issues is central to making better health decisions. This applies to the general public as well as your health practitioner. Probabilistic information:

- gives us an idea of the chance of a particular event occurring
- tells us what specific factors affect the probability of an event
- tells us whether an intervention changes the probability of an event.

Knowing the probabilities of the benefits and harms of different tests and treatments can help you and your practitioner make wise decisions that take account of your personal situation.
References

1. Doyle A. The Sign of Four 1859–1930.