Clinical Decision-Making

6

Stephen M. Downs

Learning Objectives

- Describe the basic concepts and main schools of probability.
- Use Bayes Theorem to update probabilities in the face of new evidence.
- Recognize potential biases and heuristics in probability estimation and decision making.
- Construct and analyze decision trees.
- Apply axioms of expected utility theory to quantify preferences in decision models.
- Assess trade-offs of cost and clinical outcomes using cost-effectiveness analysis.
- Identify advanced decision-modeling techniques used in CDSS.
- Explain the relationship between decision science and clinical informatics.
- Understand real-world contexts for clinical decision analysis and CDSS.

Practice Domains: Tasks, Knowledge, and Skills

The following core competencies are covered in this chapter:

- K026. Decision science (e.g. Bayes theorem, decision analysis, probability theory, utility and preference assessment, test characteristics, clinical decision support, shared decision making)

Case Vignette

You are working in the fast track (low acuity) of an urban primary care clinic. The next patient to be seen is a 34-year-old woman with a chief complaint of a sore throat. Before you enter the room, what is the probability that she has strep throat (streptococcal pharyngitis)? What questions and physical examination findings will you rely on to help narrow down the differential diagnosis? Are there any decision support tools that you could use to help you make the correct diagnosis?

Introduction

Decision-making under conditions of uncertainty is challenging. There may be many courses of action to follow, and the outcomes of those actions are not known with confidence. Although one action can lead to the most desirable result, there is a chance that it may go awry. Perhaps a safer, more middle-of-the-road approach would be better.

Consider the classic case of the patient with abdominal pain and one episode of vomiting. Her belly is moderately tender without significant rebound. Could she have appendicitis?

This is the nature of making decisions under uncertainty. Any time there are limited resources, different potential courses of action, uncertainty about what will follow the chosen action, and preferences over the potential outcomes, the benefits of formal decision-making techniques come into play.

Cognitive Aspects of Decision-Making

As a decision-making machine, the human brain is prone to errors. As recently as 1944, humans were thought of as rational agents whose thoughtful actions could explain the behavior of, for example, economic systems [1]. Decision modeling was considered descriptive of human behavior. However, by the 1960s, a growing body of psychological research showed that human decision-making could (and often did) deviate from the idealized model [2, 3]. Decision analysis moved a

S. M. Downs (✉)
Center for Biomedical Informatics, Wake Forest University School of Medicine, Winston Salem, NC, USA

Department of Pediatrics, Wake Forest University School of Medicine, Winston Salem, NC, USA
e-mail: sdowns@wakehealth.edu

J. T. Finnell, B. E. Dixon (eds.), *Clinical Informatics Study Guide*, https://doi.org/10.1007/978-3-030-93765-2_6

presumed description of decision-making to a normative prescription for how decisions should be made [4].

Probability: The Heart of Rational Decision Making

Probability estimation is a well-understood metric for representing uncertainty. But even this has been a relatively new notion in human history [5]. What is a probability? A **probability** is a number between zero and one representing the likelihood (or our belief) that something will happen or that a proposition is true. What is the probability a roll of two dice will come up with "snake eyes" (two ones)? What is the probability an infant with fever will have a urinary tract infection? What is the probability the president of the United States will walk into your office on his hands?

A probability of zero means *absolute* certainty that an event will not happen. A probability of one means *absolute* certainty that it will. All other probabilities are gradations in between. In mathematical terms, p(A) represents the probability of A. Probabilities have certain behaviors described as axioms. An **axiom** is a statement accepted as true for the purposes of developing and proving a theorem [6]. In addition to zero and one representing certainty, the axioms include that the probability of A *and* B is equal to the probability of A times the probability of B:

$$p(\text{A and B}) = p(\text{A}) \times p(\text{B}),$$

A and B are assumed to be *independent*, a notion discussed in the section under Bayes' rule. This notion is intuitive with respect to dice. If the probability of rolling a one on a single roll of one die is 1/6, then the probability of getting one's on both of two dice is 1/6 × 1/6 = 1/36.

Finally, the probability of A or B is the probability of A plus the probability of B:

$$p(\text{A or B}) = p(\text{A}) + p(\text{B}),$$

If A and B are mutually *exclusive*, meaning they can't occur at the same time. So, the probability of getting either a one or a two on the roll of a single die is the sum of the probabilities of getting each, 1/6 + 1/6 = 1/3.

There are several schools of **probability theory**. The three most common are classical, frequentist, and subjective [7].

Classical Probability Theory

The classical school refers to the early concepts of probability. These applied to games of chance and are fairly easily understood. For example, when flipping a coin, we easily understand that the probability of getting heads is 50%. If I roll a die, I interpret the chance of getting a six as one in six. A card chosen randomly from a deck of 52 cards has a one in 52 probability of being the ace of spades.

The reader would have come up with the same probabilities, or at least understand them as reasonable. But how? Few people have flipped a coin hundreds of times, carefully tracking the percentage of times the result was heads. And among those who have, a vanishingly small minority will have gotten exactly 50% heads. Yet, we understand the "true" probability of heads to be 50%. This is the classical interpretation of probability, which can be derived from understanding the underlying mechanisms. We know that the result of a coin flip can only be heads or tails (ignoring the extremely rare case where a coin may land balanced on its edge).

Moreover, we have no reason to believe that either outcome, heads or tails, is more likely than the other. Therefore, we divide our total belief in the result (100%) evenly between the two outcomes in the so-called "sample space." Heads get 50%, and tails get 50%. Likewise, if we believe a die, when rolled, is equally likely to land on any of its six sides, the probability of it landing on any given side is 1/6.

Thus, calculation of a classical probability requires no empirical data, as it is mostly analytical. Unlike frequentist probabilities (see below), it does not require infinite sets. **Classical probabilities** are objective (as we have seen) as long as there is consensus about the underlying mechanisms. However, they require knowledge of elementary events and are strongly model-bound.

Frequentist Probability Theory

Another school of probability, widely used in scientific disciplines, is the frequentist interpretation. The concept here is that the probability of a specific outcome of an experiment can be estimated by repeating the experiment N (a large number) times. The ratio of the number of times a specific outcome occurs (n) to the number of experiments performed (n/N) is an *estimate* of the probability of that outcome [8]. This conceptualization assumes the existence of some underlying "true" probability of the outcome. It posits that this true probability could be determined if we could conduct an infinite number of experiments. Since this is impossible, **frequentist probabilities** are estimates. This is why we are fond of notions like 95% confidence intervals and p-values to tell us how far we might be from the true value. Frequentist probability theory also gives rise to the "law of large numbers," the principle that the larger the number of trials, the more precise the probability estimation.

A frequentist probability requires historical data. It is empirical and cannot be derived from first principles. The frequentist school presumes a stable world because the underlying "true" probability is assumed not to change. It requires exact replication of the experiment and cannot be

applied to a unique event. Therefore, estimating the probability of success for the first manned trip to mars could not be done in a strictly frequentist way. The experiment cannot be repeated multiple times. Frequentist probabilities are never exact because infinite replication is not possible.

Subjectivist Probability Theory

The third school of probability is the subjectivist school. Subjectivist probabilities require neither data nor formal analysis, but the subjective probability school subsumes the other schools philosophically. Subjective probabilities are the most commonly estimated and used by far and are critical to the decision modeler. To illustrate a subjective probability, answer the following question: What is the probability that you will find the word "computer" on page 100 of this book. Don't look; just write down your probability, a single number. How did you choose your probability? You might have thought about the number of pages you have read so far in this book and the number of times you read the word "computer." That would be a frequentist approach. Or you might have thought I was going to "game" the system by making sure the word "computer" appears on page 100 (classical). Or you might have considered that this is a book about informatics, so most pages will mention a computer—something between classical and frequentist. **Subjective probabilities** are best thought of as a measure of belief. They may differ from person to person, but they can be applied to all conceivable uncertainties. They deny the possibility of objective probabilities. Instead, they simply represent what is going on "between your ears," a measure of your belief that the word "computer" is on page 100 [7].

Now, look at page 100. Did you find the word "computer?" So if your subjective probability was 10%, were you wrong? If it was 90%, were you wrong? No, because you were only expressing your degree of belief that "computer" was on page 100. The only way you could conceivably have been "wrong" would be if you had said the probability was zero or 100%. Now that you've looked at page 100, of course, your subjective probability has changed.

I emphasize subjective probabilities because they are the most commonly used and because their necessity is inescapable in clinical practice and formal decision modeling. Consider the physician who sees a patient with a sore throat. According to the Centor criteria [9, 10], the probability this patient has streptococcal pharyngitis can be estimated by adding points for the patient's age, signs and symptoms as follows:

- History of fever
- Tonsillar exudates
- Tender anterior cervical adenopathy
- Absence of cough
- Age <15 add 1 point
- Age >44 subtract 1 point

The probability of strep is estimated based on the score. A score of –1, 0, or 1 implies the probability of strep is <10%. If the score is 2 points, the probability of strep infection is 15%; if 3, 32%. If the score is 4 or 5, the probability is 56%. This is a purely frequentist probability estimation because it is based on the number of times strep was found in the throats of a sample of patients with different combinations of these findings. But if we learn that two other household members have had positive strep throat cultures or observe that the patient has a scarlatiniform rash—findings not included in the Centor criteria—we would certainly adjust our estimate upwards because our *belief* that the patient has strep would be increased. Now the probability is subjective. No patients or circumstances are identical to those in a randomized controlled trial or a formal observational study. So subjective adjustment of probabilities is the norm.

Subjective probability is equally indispensable in formal modeling simply because all probabilities must be represented in a formal model. There are rarely clinical studies that provide a robust and appropriate measurement of all needed probabilities.

Biases in Estimating Probability

Despite the necessity for subjective probability estimates, a large body of literature shows that humans are naturally prone to errors or biases in their probability estimates. Fortunately, there are techniques for improving one's skills at probability estimation.

The human mind uses various "tricks" to estimate probabilities. Kahneman and Tversky described the best known of these tricks in their seminal work [2, 3]. To illustrate, consider this well-known example:

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with discrimination and social justice issues and participated in antinuclear demonstrations. Please check off the most likely alternative:

❑ Linda is a bank teller.
❑ Linda is a bank teller and is active in the feminist movement.

In their study, Kahneman and Tversky found that 10% of respondents chose the first alternative and 90% chose the second, even though quick reflection will reveal that the population of bank tellers active in the feminist movement is a strict subset of all bank tellers. Therefore, Linda is at least as likely a bank teller as she is a bank teller and active in the feminist movement.

This cognitive error is known as the representativeness heuristic. A **heuristic** is a mental shortcut to solving a

problem, producing an approximate solution. The **representativeness heuristic** involves gauging the probability of an event based on how representative it seems to be of a class. In this case, a woman who was deeply concerned with issues of discrimination and social justice and participated in anti-nuclear demonstrations sounds like someone who would be active in the feminist movement. This representativeness apparently made 90% of respondents overlook the logic of the problem.

Similar problems occur with what Kahneman and Tversky call the **availability heuristic**. Like the representativeness heuristic, availability refers to estimating the likelihood of an event based on how easily it comes to mind. Although this works much of the time, it can lead one astray. For example, most people believe breast cancer is the number one killer of women because of this condition's massive press. While over ten times more women die each year from cardiovascular disease than breast cancer [11].

One variant of the availability heuristic is the **vividness effect**. This bias occurs because we tend to rate the probability of something based on how vividly it is described or, sometimes, how emotionally evocative it is. When this chapter was first written, according to surveys, Americans were nearly as worried about Ebola as they were about catching the flu. At that time, exactly one person in the US had died from Ebola—ever. *Every year*, between 3000 and 49,000 people die of influenza in the US alone. In most years, this is higher than the number who have *ever* died of Ebola *anywhere*. But we heard so much more about Ebola, sometimes in excruciating detail. It makes getting Ebola seem more real and, therefore, more likely.

Combining Probabilities: Bayes Theorem

Estimating probabilities is one thing, but the more common challenge in medical reasoning (and any other reasoning for that matter) is how to update probabilities given new evidence. Although we do it all the time (a patient suspected of having an infection has an elevated white blood count or a pedestrian judges the traffic volume before endeavoring to cross the street), we often do it badly. Test yourself.

Table 6.1 Classic 2-by-2 contingency table

		Truth (disease)		
		Positive	Negative	
Test	Positive	9	999	1008
	Negative	1	8991	8992
		10	9990	

The average patient has a one in one thousand chance of having a disease. A test for that disease has 90% sensitivity and 90% specificity (pretty good!). The test is positive. Now, what is the chance the patient has the disease? Write down your guess. In a test of Harvard medical students, most guessed it was in the neighborhood of 90% [12]. The probability is slightly less than 1%. The math required to avoid this potentially catastrophic miscalculation is surprisingly straightforward.

Let's begin with the classic 2-by-2 contingency table (Table 6.1).

The table depicts 10,000 hypothetical patients. In the columns, we see that one in one thousand, ten patients have the disease (truth), and 9990 do not. If the test is positive in 90% of those with the disease (the definition of sensitivity), then 9 of the ten patients with the disease will have a positive test result. Among the 9990 without disease, 90%, or 8991, will have a negative test (the definition of specificity). So now, if we look across the rows, we see that of all 1008 patients with a positive test, nine or about 0.9% have the disease. The rest are *false positives*. Of the 8992 patients who have a negative test, only one *false-negative* will have the disease.

Using a 2-by-2 table to make these calculations is a bit cumbersome. However, the calculations can be made in a closed-form equation. We use the term *prevalence* to refer to the probability of disease before the test is performed (also called the *prior probability*) and the term *positive predictive value* or *PPV* (also called *posterior probability*) to refer to the probability of disease after a positive test is observed. Note the *negative predictive value* or *NPV* is the *posterior probability of no disease* after observing a negative test. We can calculate the *PPV* as follows:

$$PPV = \frac{Prevalence \times Sensitivity}{Prevalence \times Sensitivity + \left(1 - Prevalence\right) \times \left(1 - Specificity\right)} \tag{6.1}$$

A more general form of this equation, using terminology introduced earlier in the chapter, is:

$$p\left(D|T\right) = \frac{p\left(D\right) \times p\left(T|D\right)}{p\left(D\right) \times p\left(T|D\right) + p\left(\neg D\right) \times p\left(T|\neg D\right)} \tag{6.2}$$

where $p(D)$ is the prior probability of disease, $p(T|D)$ is the probability of a positive test given disease (the *sensitivity*), $p(\neg D)$ is the probability of not having the disease (1-*prevalence*), and $p(T|\neg D)$ is the probability of a positive test given not disease (1-*specificity*).This is Bayes' formula,

attributed posthumously to Reverend Bayes in 1763 [12]. A more compact version of Bayes' formula can be derived by dividing formula (6.1) above by the equivalent formula for calculating the *negative predictive value* as follows:

$$\frac{PPV}{1-NPPV}=\frac{\left[\dfrac{Prevalence\times Sensitivity}{Prevalence\times Sensitivity+\left(1-Prevalence\right)\times\left(1-Specificity\right)}\right]}{\left[\dfrac{\left(1-Prevalence\right)\times\left(1-Specificity\right)}{Prevalence\times Sensitivity+\left(1-Prevalence\right)\times\left(1-Specificity\right)}\right]} \tag{6.3}$$

Formula (6.3) reduces to

$$\frac{PPV}{1-NPPV}=\frac{Prevalence}{1-Prevalence}\times\frac{Sensitivity}{1-Specificity}$$

The term $\frac{Prevalence}{1-Prevalence}$ is referred to as the *odds* of disease; it is the probability divided by one minus the probability. The term $\frac{Sensitivity}{1-Specificity}$ is known as the *positive likelihood ratio (LR⁺)*. The term $\frac{PPV}{1-NPPV}$ is the *posterior odds* of disease ($odds_{post}$). Thus, Bayes' formula can be expressed as

$$Odds_{post}=Odds_{prior}\times LR^{+}$$

The *posterior odds* following a negative test are calculated in the same way, using the *negative likelihood ratio (LR⁻)*, which is given by $\frac{1-Sensitivity}{Specificity}$.

This is known as the *odds ratio* form of Bayes' formula [13]. It can become relatively easy to use this formula to estimate posterior probabilities in one's head with practice. Let's revisit our earlier example of a patient with a one in one thousand chance of disease and a positive test with 90% *sensitivity* and 90% *specificity*. The *prior probability* of disease is one in a thousand, so the *odds* of disease is (1/1000)/(1 – 1/1000), which is very close to 1/1000. (For very low probabilities, the odds are approximately equal to the probability.) The *positive likelihood ratio* is the *sensitivity* divided by one minus *specificity* or .9/.1=9. The *posterior odds* are nine times 1/1000 or 9 in 1000. The *posterior probability* is the *odds/(1+odds)* or 0.009/(1+0.009), which is very close to 0.009, or 0.9%, as we saw with the 2-by-2 table above.

Table 6.2 Sample collection of likelihood ratios (LR) for a hypothetical decision support system. Each LR describes the relationship between evidence (symptoms, findings, test results) and a given diagnosis (see text)

Evidence	LR^{+}	LR^{-}
Symptom A	2.3	0.8
Exam Finding B	3.0	0.2
Test Result C	4.1	0.85
Test Result D	3.1	0.1

Bayes' formula's odds ratio invites an attractive algorithm for computing updated probabilities as new evidence is acquired. Because we can treat the posterior odds of disease following one test as the prior odds of disease for a subsequent test, we can string together likelihood ratios to calculate the posterior odds after an arbitrary number of bits of evidence have been evaluated. What's required is a *prior probability* of disease and a catalog of positive and negative *likelihood ratios* for the evidence to be considered (Table 6.2)

A diagnostic program could evaluate the likelihood of a diagnosis with a prevalence of 2% in a patient who has

$$Odds_{prior} := Prevalence/(1\text{-}Prevalence) = 0.02/0.98 = 0.0204$$

$$Odds_{post} := Odds_{prior}\times LR^{+}_{A}\times LR^{+}_{B}\times LR^{-}_{C} = 0.0204\times 2.3\times 3.0\times 0.85 = 0.12$$

symptom A, exam finding B, and negative test C, but for whom the results of test D are unknown as follows:

$$\text{PPV} := \text{Odds}_{\text{post}}/(1 + \text{Odds}_{\text{post}}) = 0.11, \text{ or } 11\%$$

With a sufficient knowledge base of LRs, such a diagnostic program could process an arbitrary number of findings, returning an updated probability each time. However, there is one critically important caveat. The relationship of each finding to the hypothesized diagnosis must be *conditionally independent* of the other findings. In other words, the probability of exam finding B, given the diagnosis, must not depend on the presence or absence of symptom A. This assumption is rarely precisely true. However, it is often close enough that the algorithm works. This approach has been successfully employed in several decision support systems [14–16].

So far, we have only considered Bayes' formula for the binary case in which two hypotheses are being considered, i.e., that the patient has the disease or the patient does not have the disease. The formula is much more general and can consider an arbitrary number of mutually exclusive and exhaustive hypotheses. The posterior probability of a given hypothesis, H_1, is given by the formula

$$p(H_1|E) = \frac{p(H_1) \times p(E|H_1)}{\sum_{i=1}^{N} p(H_i) \times p(E|H_i)}$$

The posterior probabilities for the other hypotheses H_2 through H_N are calculated in the same fashion. Although this formulation is not as compact as the odds ratio form, complex diagnostic problems can be addressed with an adequate knowledge base of conditional probabilities. Likelihood ratios can also be expanded to multiple levels of a test result (interval likelihood ratios) to account, for example, for how a 3+ leukocyte esterase test result increases the probability of urinary tract infection more than a 1+ result [17].

Decision Science

Decision analysis (DA) is a method for choosing a course of action under conditions of uncertainty. For the purposes of DA, a decision can be thought of as having three components

1. Two or more alternative courses of action,
2. Uncertainty about the outcomes of those courses of action, and
3. Preferences for the different outcomes that are possible.

A decision also involves an irreversible commitment of resources (no "do-overs").

DA provides a formalism for representing each of these components.

1. Courses of action (and their potential consequences) are represented in a decision model, often a decision tree as discussed below.
2. Uncertainty is represented with probabilities and Bayes' theorem, as we have discussed in the previous section.
3. Preferences are represented with utilities, a numeric quantification of an individual's relative preferences for different outcomes. These are discussed in the next section.

Decision Trees

A **decision tree** is a branching diagram representing courses of action that can be taken and the events that may happen as a result. Consider the following example. A 12-year-old patient presents to an emergency room with a mild fever and abdominal pain. She has vomited once. Based on a detailed history and physical examination, you have decided that there is a 30% chance she has appendicitis. You have decided on two possible courses of action. You can take her directly to surgery and remove her appendix. This surgery comes with a small risk of surgical death, about 1 in 10,000. Alternatively, you can observe her in an observation unit overnight. Let's make some simplifying assumptions. First, assume that if she *doesn't* have appendicitis, she has a self-limited viral infection, and if you observe her overnight, she will recover and go home.

On the other hand, if she *has* appendicitis and you choose to observe, there is a 35% chance that her appendix will rupture. In that case, she will require surgery, and the risk of surgical death is ten times higher. If her appendix does not rupture, she will still need surgery (because she has appendicitis), but the risk of death will not be higher.

Figure 6.1 shows a decision tree representing this situation. The tree consists of a series of nodes with branches coming out of them. It is read from left to right. There are three types of nodes; the square node on the left is a *decision node*. The branches coming from a decision node represent the choices under the decision maker's control, in this case, taking the patient to surgery or observing overnight. Each of these branches leads to a round *chance node*. Each branch coming from a chance node represents something that might or might not happen but over which the decision-maker has no direct control. The branches are associated with probabilities. In the case of the "Surgery" node, the chance of "Surgical Death" is 0.0001 (one in ten thousand). The chance of "Survive Surgery" is 0.9999. In statistical vernacular, chance nodes represent random variables, with the branches representing possible values in the outcome space. As such, the branches must be mutually exclusive and exhaustive, meaning the probabilities of the branches emanating from a given chance node must sum to 1.0.

The third type of node is a *terminal* or *value* node, shown along the right side of Fig. 6.1. These nodes hold numeric

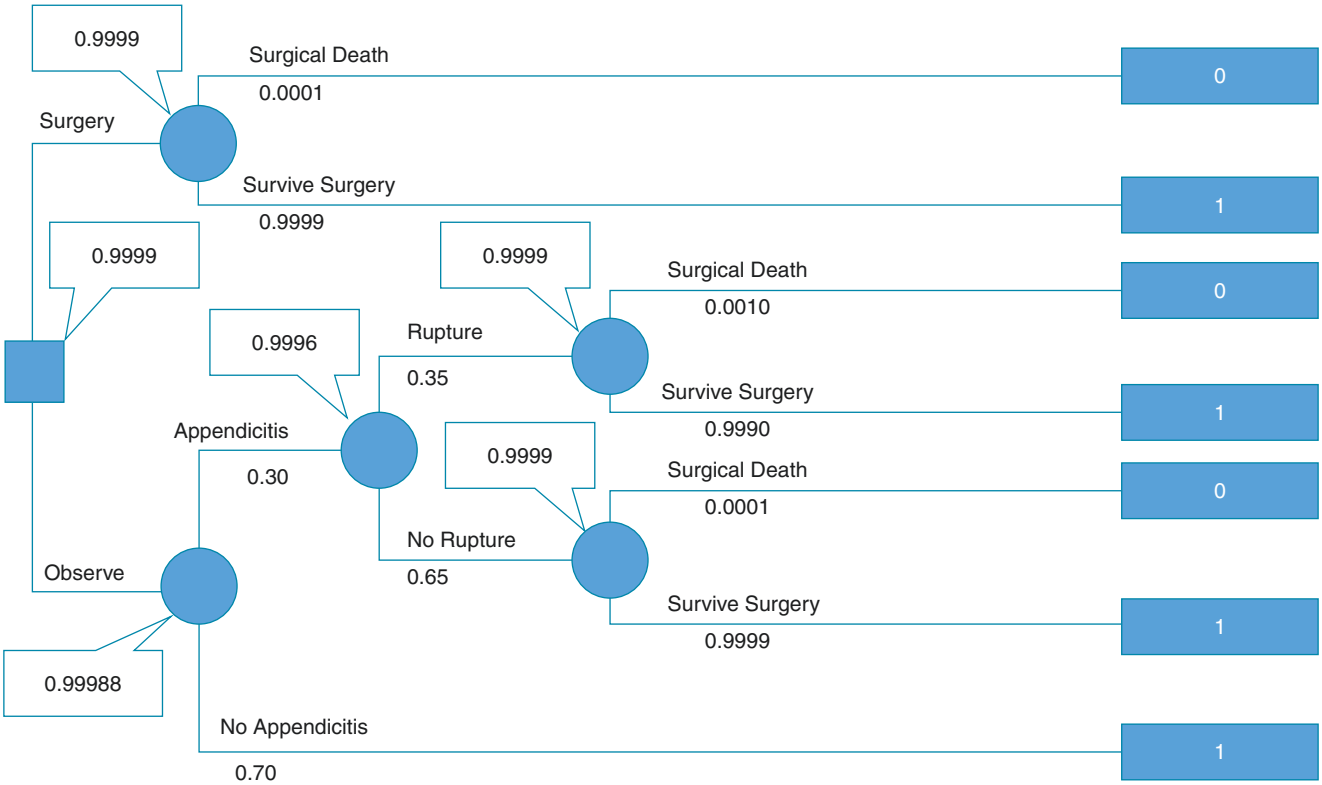

Fig. 6.1 The appendicitis decision tree. As described in the text, this decision tree illustrates the three main types of nodes in a decision tree: square decision nodes, round chance nodes, and terminal nodes at the end of each path

representations of the decision-maker's values on the outcomes at the end of the decision tree. This numeric representation is called a utility. For the moment, we will use the world's simplest utility measure, 1 for surviving and 0 for dying. The theoretical basis for assigning more precise values to outcomes is discussed in the section "Expected Utility Theory" below.

Following the tree from left to right, if the decision-maker decides on the surgery option, we have said there is a 9999 in 10,000 chance the patient will survive. If observation is chosen, there is a 30% chance the patient will have appendicitis. In that case, there is a 35% chance the appendix will rupture. If the appendix ruptures, there is a one in 1000 (0.001) chance of surgical death and a 999 in 1000 chance of surviving an appendectomy. If the appendix does not rupture, the chance of surgical death from an appendectomy is still 0.0001. Finally, if the patient does not have appendicitis, her symptoms resolve, and she goes home.

The decision tree is analyzed moving from right to left, using a recursive algorithm. If a node is a utility node, its value is its utility. If it is a chance node, its value is the expected value of its branches, that is, the sum across its branches of the product of the value of the branch times the probability of the branch. If the node is a decision node, its value becomes the value of whichever of its branches has the highest value—the decision that should be taken.

The values of the nodes in Fig. 6.1 are shown as bubbles pointing to the nodes. The expected value (EV) of the *Surgery* node is the value of dying times the probability of dying plus the value of surviving times the probability of surviving, $(1 \times 0.9999) + (0 \times 0.0001) = 0.9999$. The value of the *Rupture* node is $(1 \times 0.9990) + (0 \times 0.0010) = 0.9990$. The value of the *Appendicitis* node is $(0.35 \times 0.9990) + (0.65 \times 0.9999) = 0.9996$. Finally, the value of the *Observe* node is $(0.30 \times 0.9996) + (0.70 \times 1) = 0.99988$. Because the EV of *Observe* is lower than EV of *Surgery*, surgery is the preferred option.

The thoughtful reader will have some objections to this simple analysis. First, the difference in the EVs of the surgery and observation options seems trivially small, only two in 100,000. This decision seems like a "close call" that may change with minor changes in our estimates of probabilities and utilities. This is a legitimate complaint that we will address in the section "Sensitivity Analysis" below. A second concern might be that our utilities, 1 for survival and 0 for death, maybe overly simplistic. Surely, a patient would rather be observed overnight and go home than have a ruptured appendix and undergo emergent appendectomy and treatment for peritonitis. A more nuanced approach to quantifying preference is discussed in the section "Expected Utility Theory."

A third point might be that we have missed an alternative. Instead of choosing surgery or observation, perhaps we can perform a test that will help us decide. The option of using a diagnostic test is easily modeled with a third branch from the decision node, as shown in Fig. 6.2. We have modeled a test with 70% sensitivity and 80%

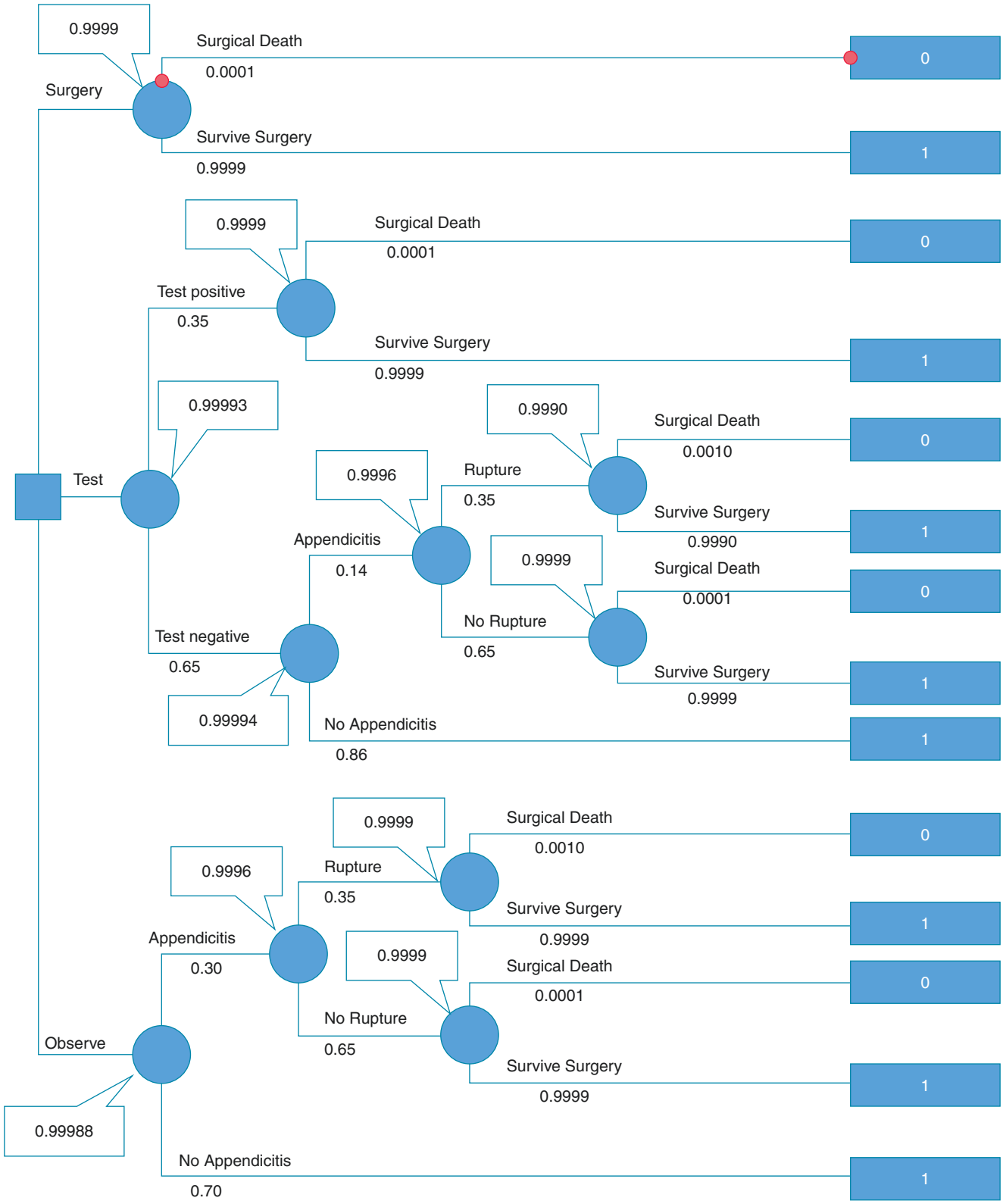

Fig. 6.2 The appendicitis decision tree with a "test" node. As described in the text, this version of the appendicitis decision tree includes the option of obtaining a test to decide how to treat the patient

specificity. Between the *Surgery* and *Observation* nodes, we have inserted a *Test* node. Under the assumption that we would take the patient to surgery if the test is positive and observe the patient if negative, the *Test Positive* branch has the same structure as the *Surgery* branch. The *Test Negative* branch has the same structure as the *Observe* branch, assuming that we will respond to a positive or negative test, respectively.

However, note that the probability of appendicitis given a negative test is now 14% instead of 30%. This 14% is calculated using Bayes' theorem, the probability of disease given a negative test or one minus the negative predictive value (see above). The probability of a positive test is given by $p(T^+|D) \times p(D) + p(T^+|\neg D) \times p(\neg D)$, the denominator of Bayes' theorem (Eq. 6.2 above).

We calculate the expected utility of the *Test* node in exactly the same way we did for the other two branches, getting a value of 0.99993, slightly higher than the EV of surgery. So the test option is the best. The difference in expected value between the best option without the test (surgery at 0.99990) and the expected value of testing (0.99993) is known as the *expected value of information* from the test.

But now let us consider another scenario, another patient with abdominal pain, but with higher fever, vomiting, and pain that is more typical for appendicitis, with migration to McBurney's point. Your subjective judgment is that the patient has a 50% chance of having appendicitis. When we evaluate the tree, the results are those in Fig. 6.3. Some find it surprising that the EV of testing has fallen below the EV of surgery. In other words, it is worse to obtain more information with the test than to just take the patient to the operating room. The test offers no *value of information* in this scenario.

To understand why this is so, consider the six probabilities that have changed, circled in Fig. 6.3. The probability of a positive test has gone up to 45%, and the probability of a negative test has gone down to 55%. More importantly, the probability of appendicitis given a negative test (the false-negative rate) has increased to 27%. In other words, if the test is negative (and we choose to observe), there is still a 27% chance the patient has appendicitis. Which decision is best depends on the prior probability of appendicitis.

Sensitivity Analysis

The exercise of varying a parameter in a decision model (like the prior probability of appendicitis) to see how it effects the decision is known as **sensitivity analysis**. Figure 6.4 shows a one-way sensitivity analysis of the probability of appendicitis. The x-axis shows the probability of appendicitis varied from 0 to 100%. The y-axis shows the expected value. Each line on the graph represents one of the three strategies—surgery, test, observe.

When the probability of appendicitis is low, *Observe* has the highest EV. As the probability of appendicitis goes up, the EV of *Observe* drops rapidly while the EV of S*urgery* stays the same (because the risks of surgery are the same regardless of the probability of appendicitis). The EV of *Test* drops more slowly as the probability of appendicitis rises. We see that at low probabilities, *Observe* is best. At high probabilities, *Surgery* is best. Only in the middle area does *Test* have the highest EV. The points where the lines cross are known as thresholds, and they represent points where the best decision changes. Figure 6.4 has dotted lines projecting the thresholds onto a "threshold bar" at the bottom [18]. This bar represents a decision rule suggesting which option is best given the estimated risk of appendicitis.

Expected Utility Theory

One objection to our appendicitis decision tree is the way the outcomes are valued. All outcomes resulting in survival were counted as 1, and those resulting in death were counted as 0. However, spending a night in observation with no surgery, is certainly better than surviving after having a ruptured appendix, requiring emergency surgery and resulting in peritonitis-although both result in survival. A more nuanced measure of preference is needed. That measure is known as a utility, and we describe the theory behind it here.

To develop the theory, let's consider a decision with a more quantifiable outcome, money. Imagine that you have the opportunity to play a game. In the game, a coin will be flipped. If the coin comes up heads, you will win $20. If it comes up tails, you win nothing. You have to pay to play this game. So there is a choice: pay to play or keep your money. Stop now and ask yourself what's the most you would pay to play this game. To help make this decision, you might calculate the EV of the game and compare it to the cost of playing. Assuming a "fair" coin, the EV of the game is 50% times $20 plus 50% times $0, or $10. If you are happy with this result, you should be willing to pay anything up to $10 to play the game because the EV of the game worth the same as $10 in your pocket. However, many years of experience (and research) have shown that the vast majority of people are unwilling to pay anything close to $10 for this game. How about you? This unwillingness to pay an amount for a gamble equal to the EV of the gamble has been termed risk aversion.

So perhaps the whole EV idea doesn't work. Nicolas Bernoulli came up with an even more dramatic example[19]. Imagine a game in which we will flip a coin. If it lands on heads, you win two dollars. If it lands on tails, the game

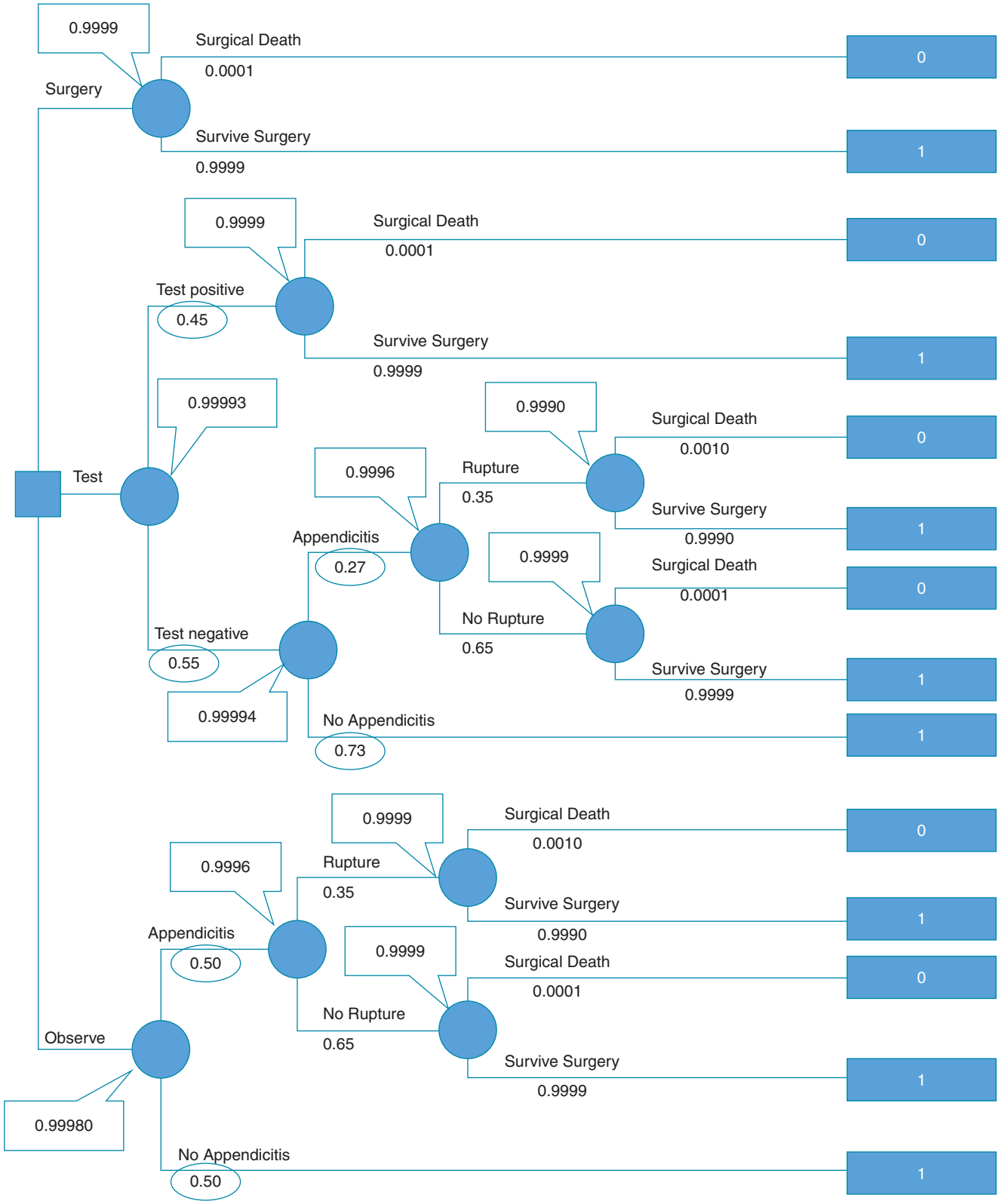

Fig. 6.3 The appendicitis decision tree with the prior probability of appendicitis increased to 50%, illustrating that which option is best changes as the parameters in the decision model change. The circled probabilities are those that change as the prior probability of appendicitis is increased

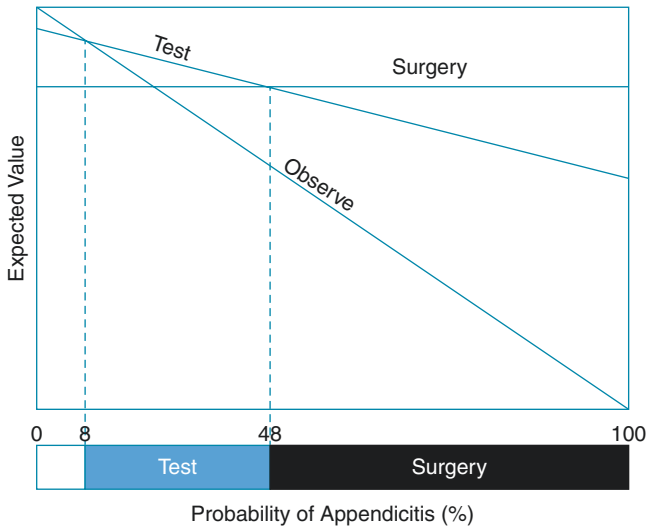

Fig. 6.4 One way sensitivity analysis of the prior probability of appendicitis. The x-axis shows the probability of appendicitis. The y-axis shows the expected value of each decision option as the probability increases. Points where the lines cross are known as thresholds

ends. Otherwise, we flip again. If you get a second heads, you win $4; a third, $8; a fourth, $16; and so forth, doubling each time you get heads but ending as soon as you get tails. How much would you pay to play that game? Most people would pay a few dollars at most, but the EV of this game is infinite because the infinite series, $\lim_{n\to\infty}\Sigma\left(\frac{1}{2^n}\right)\times 2^n$, is unbounded.

Nicolas Bernoulli appears to have contradicted EV as a basis for decision-making. However, his cousin, Daniel Bernoulli, proposed a solution, suggesting that the marginal benefit of each unit of money gained decreases as the person receiving it gains more and more. To paraphrase Bernoulli, a dollar surely means more to a pauper than to a rich man.

This idea implies that we need a new metric, a function on dollars that behaves the way we want it to behave—that is, its expected value is a basis for making a decision. Such a function is known as a **utility**. **Expected utility theory** was first formalized by von Neumann (a mathematician) and Morgenstern (an economist) in 1944 [1]. Starting with a set of axioms or postulates, they developed a formal proof that the expected value of their utility function should be the basis of rational choice. Raiffa and Howard have developed more intuitive versions of this proof [20]. What follows is adapted from Howard's axioms of expected utility theory.

The axioms of expected utility theory, as framed by Howard, are (1) orderability, (2) transitivity, (3) monotonicity, (4) decomposability, (5) continuity, and substitutability [21]. To illustrate how they lead to utility theory, imagine you have a condition called the *clinical epidemioma (CE)*. Left untreated, a CE is uniformly and rapidly fatal. Of course, CE is not a real disease; I have invented it for this illustration. There are three treatments available: (1) Tumorex, which results in a 10-year survival in the 50% of patients whose bodies absorb it; (2) GastroSorb, which is absorbed by all patients but is effective in 50% of tumors, resulting in 10-year survival; and (3) Mediocrin, a generic that results in 4-year survival for all patients who take it. In one arm of a randomized controlled trial, the combination of Tumorex and GastoSorb was tried. The combination was fatal in 20% of patients because of an enzyme in 40% of patients that renders GastroSorb toxic in the presence of Tumorex.

Figure 6.5 illustrates the choice of treatments in the CE in a decision tree.

At first glance, the combination seems like the obvious winner because it offers the highest life expectancy (5.5 years), but let's review the axioms of expected utility and see how they apply.

1. **Orderability** simply means that we are willing to order the outcomes in our decision problem according to preference. Two outcomes may be deemed equally desirable. In the CE example, we probably would prefer 10 years to 4 years to 0 years.
2. **Transitivity** says that if we like A better than B and B better than C, then we must like A better than C. A violation of this axiom can turn you into a "money pump" because, if it is not true, I can get you to pay me a small amount to take B in exchange for C, then a bit more to take A in exchange for B. But then I can get a bit more to take C in exchange for A and continue like this indefinitely.
3. **Monotonicity** means that, given two gambles with prizes A and B, if I like A better than B, I will prefer the gamble that gives me the higher probability of A–I want the gamble with the higher probability of the thing I like better.
4. **Decomposability** is also known as the "no fun in gambling" axiom. It states that all we care about is the probabilities of the outcomes, not how the sequence of events leads to them. For example, Tomorex is 50% absorbed but 100% effective, and GastroSorb is 100% absorbed but 50% effective. These are equivalent because both represent a 50% chance at the outcome, 10 years.
5. **Continuity and substitutability** states that for any three outcomes (for example, 0, 4, and 10 years), there exists some probability, p, at which the decision-maker is indifferent between a lottery with probability p of the best outcome and 1-p of the worst outcome and taking the intermediate outcome with certainty. In the case of the CE, given a choice between 4 years for sure and a gamble with a probability, p, of living 10 years and a probability, 1-p, of dying, there is some probability, p, at which the certainty and the lottery would have equal preference.

Let's consider just the Combination branch to show how these can be applied to the CE tree in Fig. 6.5. The

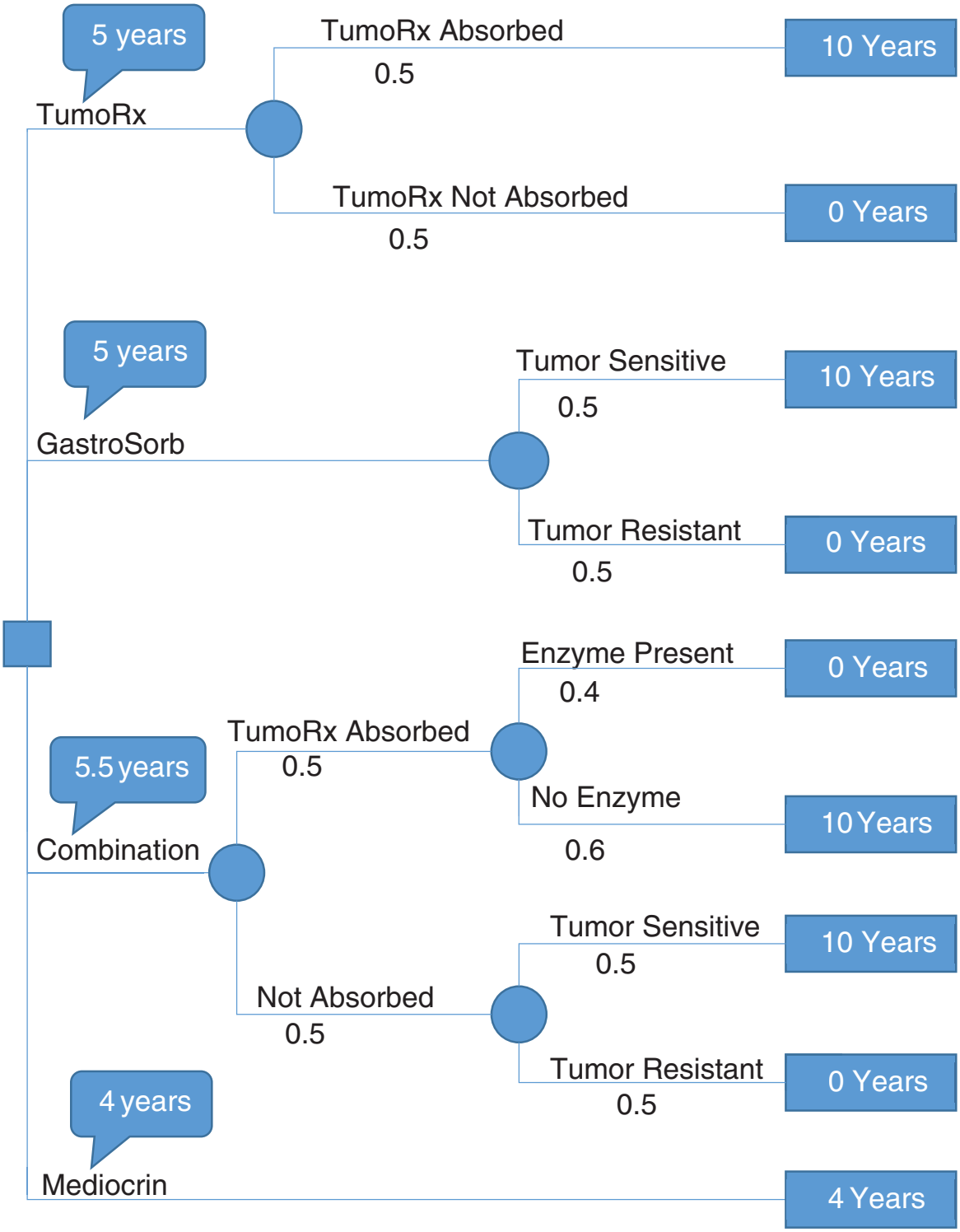

Fig. 6.5 Decision tree illustrating the choice of treatments for the *clinical epidemioma*. The combination treatment appears to offer the highest expected survival. However, application of the axioms of expected utility theory shows that this may not be the best choice (see text)

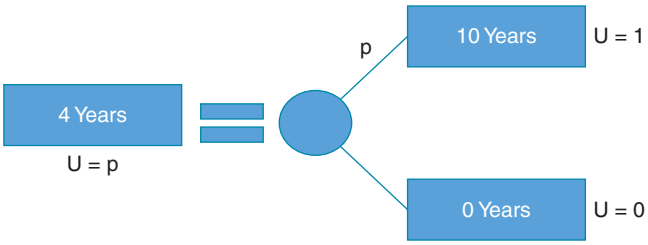

Fig. 6.6 The standard gamble. The relative utility values for outcomes in a decision analysis are calculated in threes. A forced choice is set up between a gamble, consisting of a probability, p, of the most preferred outcome and probability, 1-p, of the least preferred outcome, or a certainty of the intermediate outcome. The probabilities are adjusted until the decision maker is indifferent between the gamble and the certainty. At this point, the utility of the certainty is equal to the expected utility of the gamble. If the utility of the most preferred outcome is set to 1, and the utility of the least preferred set to 0, the utility of the certainty is equal to p

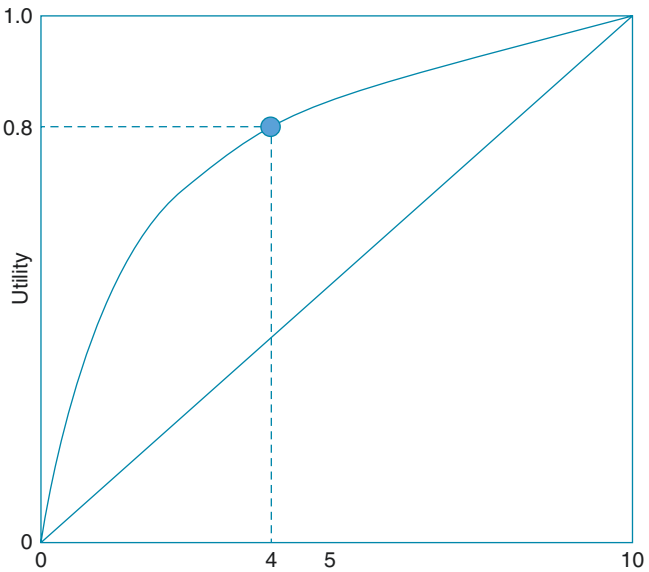

Fig. 6.7 Utility curve on years of life. The curve shows one decision maker's utilities on remaining years of life as a function of years of life. The figure highlights the that the utility of 4 years of life, U(4 years) is 0.8 on a scale where U(0 years) = 0 and U(10 years) = 1. The curve is bowed up and to the left (concave up), indicating the decision maker is risk averse

decomposability axiom says that multiplying and adding can change that branch to a single gamble with a 55% chance of 10 years and a 45% chance of 0 years without changing our preference for that option. The continuity and substitutability axiom says that, in the *Mediocrin* branch, we can replace the 4 years for sure with a gamble between 10 years at probability, p (where p is the indifference probability), and 0 years with probability 1-p without changing our preferences.

Comparing the *Combination* and *Mediocrin* branches, we compare two gambles with the outcomes 10 and 0 years. One offers 10 years with a probability of 55% and the other a probability p. So the preferred option depends on the indifference point, p. This is assessed using the standard gamble (or standard reference gamble described below).

The Standard Gamble

Von Neumann-Morgenstern (vNM) utilities are assessed with the **standard gamble**. This is simply a process for finding the indifference point. This is done by setting up a trade-off between a gamble with the best and worst outcomes and an intermediate outcome for certain, as illustrated below (Fig. 6.6). A series of forced-choice questions are asked as follows. A value between 0 and 1 is assigned to p (e.g., 50%), and the respondent (decision maker) is asked whether she would prefer a gamble with a 50% chance of 10 years (the best outcome) and a 50% chance of 0 years (the worst outcome), or if she would rather have 4 years for sure, referred to as the certain equivalent. If she says she would prefer 4 years for sure, p is adjusted upward, perhaps to 75%. Then the respondent is asked whether she would prefer a gamble with a 75% chance of 10 years and a 25% chance of 0 year, or if she would rather have 4 years for sure.

The probability, p, is adjusted in this way until p has a value at which the respondent cannot choose between the alternatives. For the standard gamble in Fig. 6.6, a common indifference point is at about p = 80%. For convenience, we arbitrarily set the utility of the best outcome in a decision to 1 and the utility of the worst outcome to 0. Thus, at the indifference point, the value of the intermediate outcome is the expected utility of the gamble or p. If the respondent were indifferent at an 80% probability of 10 years (and a 20% risk of death), the utility of 4 years (the certain equivalent) would be 0.8.

This process can be repeated for all of the outcomes in a decision tree with preference weightings between the best and the worst. And the proof put forth by von Neumann and Morgenstern means that the expected utility is an appropriate basis for choosing alternatives. If these utility values are plotted as a function of the outcomes, the result is typically a curve, as shown in Fig. 6.7. This curve, said to be concave up, is typical of risk aversion. It is consistent with Daniel Bernoulli's proposal that the marginal gain of each unit of outcome goes down as the total number of units goes up.

Most individuals will be risk-averse under most circumstances, but there are risk-seeking individuals and situations in which individuals will exhibit both risk-seeking and risk-averse preferences [2, 3].

Time Trade-Off

By virtue of arising from vNM expected utility theory, the standard gamble is generally considered the gold standard for utility assessment. However, because it can pose a

cognitive burden, other methods have been developed. The most important of these is the **time trade-off** (TTO) [22]. The TTO is most suitable for assessing utilities for time spent in a chronic health state. In the TTO, the respondent (decision maker) is presented with his remaining life in a chronic, less than ideal health state. For example, living with total blindness for 20 years (followed by death). He is then asked how many of those 20 years he would give up to have his vision back. This can, and often is, posed as a series of forced-choice responses. For example, would you give up 10 of those years to have your vision back? This would be repeated, adjusting the number of years in good health until an indifference point is reached, much as is done with the standard gamble.

So if the respondent is indifferent between living 20 years with blindness and living only 15 years with vision, his utility for blindness is calculated as the number of years with vision divided by the number of years with blindness, 15/20 = 0.75. Utilities derived from the TTO can be shown to be consistent with those derived by standard gamble under the assumption that the respondent is risk-neutral, something that we've said is rarely true [23]. Additionally, the TTO assumes a constant proportional tradeoff, meaning that if the trade-off were based on 10 years in a health state or 30 years in a health state, the response would yield the same ratio of ¾ described above.

Quality Adjusted Life Years

Over the last two decades, quality-adjusted life years (QALY) has become the most widely accepted utility model in medicine [24]. QALY is a multi-attribute utility model, meaning that it takes separate measures of health outcomes and combines them to form one utility measure [25]. One dimension of the QALY is the length of life measured in years. The second dimension is the quality of life during those years. Typically, but not always, the quality term is a utility, often assessed with the TTO method. Other utilities for quality adjustment can come from standardized utility indices such as the Health Utilities Index (HUI) or the EQ-5D, EuroQual [26, 27]. Utilities used to adjust QALYs must be anchored at zero for death and 1.0 for perfect health. The basic formula for a QALY is the length of life multiplied by one or more quality adjustments.

Because QALYs are normalized to 1 QALY for a year in perfect health and zero QALYs for death, QALYs for time spent in different health states can be added together to total the QALYs over changing health states even for an entire lifetime. This is especially useful for Markov models and simulations, as described below.

Cost-Effectiveness and Cost-Utility Analysis

The concept of **cost-effectiveness analysis** arises because it can be helpful to consider costs and health outcomes of a decision problem separately. As we have seen, it is possible to measure utilities for monetary outcomes and clinical outcomes. Moreover, vNM utilities can be assessed over global outcomes that include both health and monetary components. However, when different parties (e.g., government or insurance companies) are paying for health outcomes experienced by others, it can be helpful to consider cost and health outcomes separately.

This is done easily enough by assigning both a health outcome and a monetary outcome to each terminal node of a decision tree and solving the tree twice, once for each of the outcomes. The general term for this is a cost-effectiveness analysis (CEA). When the health outcome is a utility, we use the more specific term, cost-utility analysis. To illustrate, below (Fig. 6.8) is a tree for evaluating a hypothetical vaccine. The tree shows two options: provide the vaccine or don't. The tree models a probability of infection, p(inf), for the *No Vaccine* branch. The probability of infection for the *Vaccine* branch is reduced by multiplying p(inf) times one minus the vaccine's effectiveness. The terminal nodes show two values separated by a "/". The first is the cost accumulated along the path leading to the node, e.g., the cost of the vaccine + infection + hospitalization. The second is the utility, in QALYs, for that outcome. (The probabilities are not shown.)

The average or expected cost and QALYs for each alternative are shown in the corresponding bubble. The vaccine strategy costs more ($28 vs. $16) but results in a greater number of QALYs (29.98 vs. 29.97). These differences are typically examined using a marginal or incremental cost-effectiveness table, as shown in Table 6.3.

To construct Table 6.3, the strategies are listed in the first column in increasing order of cost. The average (expected) cost of each strategy is entered in the second column. The third column is the incremental cost, the difference between the cost of each strategy and the next cheapest strategy (the one above it). The average effect is entered next, followed by the incremental effect, the difference in effect between each strategy and the strategy above it. An average cost-effectiveness ratio, the ratio of the average cost to the average effect, is next. It is important to know that this number has very little meaning in isolation. *CEA must always be done in comparison between two or more competing strategies*. The last column is the incremental cost-effectiveness ratio (ICER). This is the ratio of the incremental cost divided by the incremental effect.

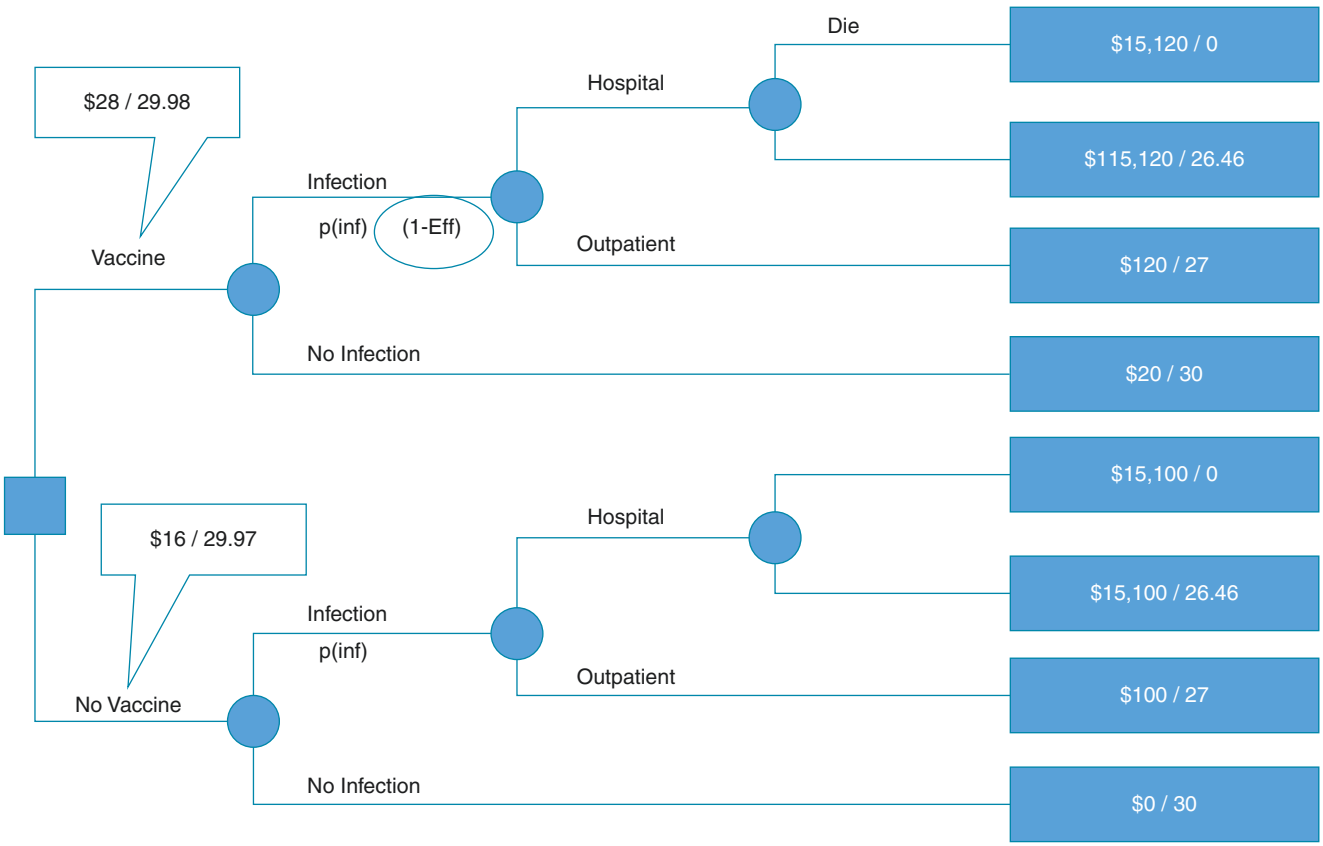

Fig. 6.8 A decision tree for conducting a cost-effectiveness analysis. The terminal nodes show a value and a cost term. The tree is solved once, calculating the expected value of each option, and a second time, calculating the expected cost of each option. The difference in cost between two options divided by the difference in value is the incremental cost-effectiveness (see Table 6.3)

Table 6.3 Table showing the calculation of incremental cost-effectiveness. The options are listed in ascending order of cost. The difference in cost and the difference in effect between the sequential options is entered. The incremental cost-effectiveness ratio is the ratio between the difference in cost and the difference in effect

Strategy	Average cost	Incremental cost	Average effect (QALY)	Incremental effect (QALY)	Cost/ effect	Incremental cost effectiveness ratio (ICER)
No vaccine	$16		29.9668		$1	
Vaccine	$28	$12	29.9834	0.0166	$1	$723

In this case, the ICER is $723. That is, the *Vaccine* strategy will cost $723 for each QALY saved. This is a very favorable ratio. Interventions with an ICER of $50,000–$100,000 per QALY are often considered cost-effective. ICERs are especially useful for comparing alternative health interventions to achieve the most efficient use of healthcare dollars [28].

Calculating Costs

We've discussed the assessment or calculation of utilities. There are some caveats to calculating costs. The first is to understand that healthcare charges rarely reflect costs. Charges are driven more by market forces than actual costs to the system. To make matters worse, healthcare systems may shift costs from one segment of care to another. Payments by government or private insurers may be closer to costs but are largely driven by negotiations between payers and providers. Payments may be appropriate measures of cost if the analysis is being done from the payer's perspective.

But perspective is all-important. Different costs and outcomes are important to payers, providers, and patients. It has been recommended that cost-utility analysis be done from a "societal perspective," accounting for all costs and health

outcomes. Still, it must be acknowledged that no one has a societal perspective [24].

It may be that the best way to calculate costs is with a cost accounting approach, which considers each of the resources that goes into delivering care as well as other costs (e.g., travel or lost work) that may be induced by an intervention or disease process.

Advanced Decision Modeling

Up to this point, we have only considered decision trees to model decision problems. However, two additional modeling approaches deserve attention, especially because modern computer technology makes them useful for computer-based decision support systems. These techniques are Markov models and influence diagrams.

Markov Models

In DA, Markov models are often used to model health states that change over time. Consider, for example, a decision regarding the choice of therapies for cancer. Following the therapy, 90% of patients enter remission and may follow any of a wide number of pathways subsequently. Each year, the patient may remain in remission or may experience a recurrence. If there is no recurrence in the first year, there may be one in the second or the third year, etc. If a recurrence does occur, it may lead to death in the first year, or the patient may spend two or more years in a chronic recurrent cancer state. To try to model all of these possible outcomes in a decision tree would be untenable.

Markov models provide a more compact method for evaluating such models. Figure 6.9 shows a simple Markov model representing this situation. Each node in the model (Well, Cancer, Dead) represents a health state. The arrows show transitions that can happen with each Markov cycle. Each transition is associated with a probability that the transition will happen in a given cycle. Each health state has an associate utility, representing the quality adjustment for the time spent in that health state.

In most computer models of Markov chains like this, it is possible to represent transition probabilities with formulas or lookup tables to make the models more dynamic.

To analyze a Markov model, we simply distribute a hypothetical cohort of patients into each of the health states and begin to simulate what happens. Table 6.4 shows how utilities, in the form of QALYs, accumulate with the first two cycles of the model.

At the initiation of the cycle, we determined that 90% of patients were in remission (the well state), and 10% had resid-

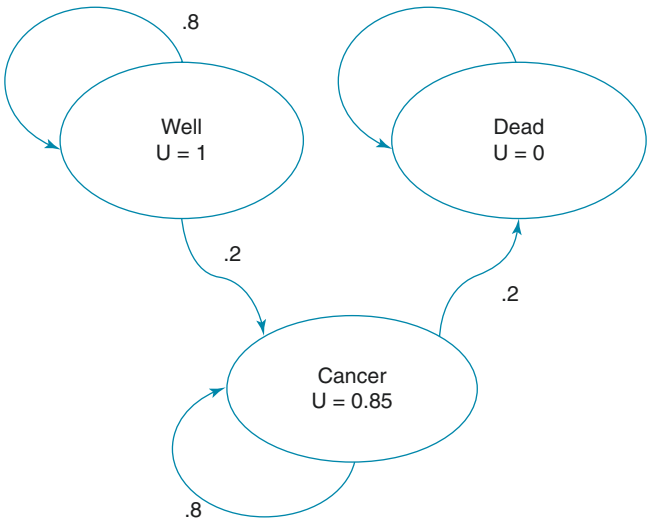

Fig. 6.9 A simple Markov model. The Markov model shows three health states, *well, cancer,* and *dead*. Arcs (arrows) between the health states represent the probability of transitioning from one health state to the next during a *Markov cycle* (for example, a year). Utility is accumulated for each cycle (see Table 6.4)

ual cancer. So in cycle 1, patients in the *Well* state each got a utility of 1. So they accrued 0.9 QALY. The 10% in the *Cancer* state had a utility of 0.85, accruing 0.085 QALY. So at the end of cycle 1, the model accumulated a total of 0.99 QALY.

In cycle 2, 80% of the patients in the *Well* state during cycle 1 remain there in cycle 2, meaning 72% are in the *Well* state for cycle 2. They have a quality adjustment of 1, so they accrue 0.72 QALY. The *Cancer* state acquired 20% of those in the *Well* state in cycle 1 and retained 80% of those in the *Cancer* state in cycle 1 for a total of 0.26 of the cohort. Their quality adjustment is 0.85 so they accrue $0.26 \times 0.85 = 0.22$ QALY. The *Dead* state acquired 20% of those in the *Cancer* state in cycle 1, but since the quality adjustment is 0, they accumulate no QALYs.

So during cycle 2, the health states accumulate a total of $0.72 + 0.22 = 0.94$ QALY. This is added to the 0.99 QALY accrued in cycle 1 to make 1.9 QALYs accumulated by the whole cohort at the end of the second cycle. This process is repeated for as many cycles as we want to model the process or until the entire cohort is in the *Dead* state and can no longer accumulate QALYs.

Influence Diagrams

An influence diagram alternative to a decision tree emphasizes the probabilistic relationships among variables [29, 30]. An influence diagram is an acyclic directed graph with three types of nodes (much like trees): decision nodes, chance nodes, and one value node. Figure 6.10 illustrates a rather generic

Table 6.4 Showing the accumulation of expected utilities (as quality adjusted life years) during two cycles of a Markov model. During each cycle, the probability of being in a state is multiplied by the utility of a cycle in that state. These are summed across states to calculate the expected utility for the cycle. This is repeated for subsequent cycles, accumulating the total expected utility for the whole simulation

Cycle	State	Probability	Expected utility	Cumulative utility
1	Well	.9	.9 × 1 = .9	
	Cancer	.1	.1 × .85 = .085	
	Dead	0	0	.99
2	Well	.9 × .8 = .72	.72 × 1 = .72	
	Cancer	(.9 × .2) + (.1 × .8) = .26	.26 × .85 = .22	
	Dead	.1 × .2 = .02	.02 × 0 = 0	.94 + .99 – 1.9

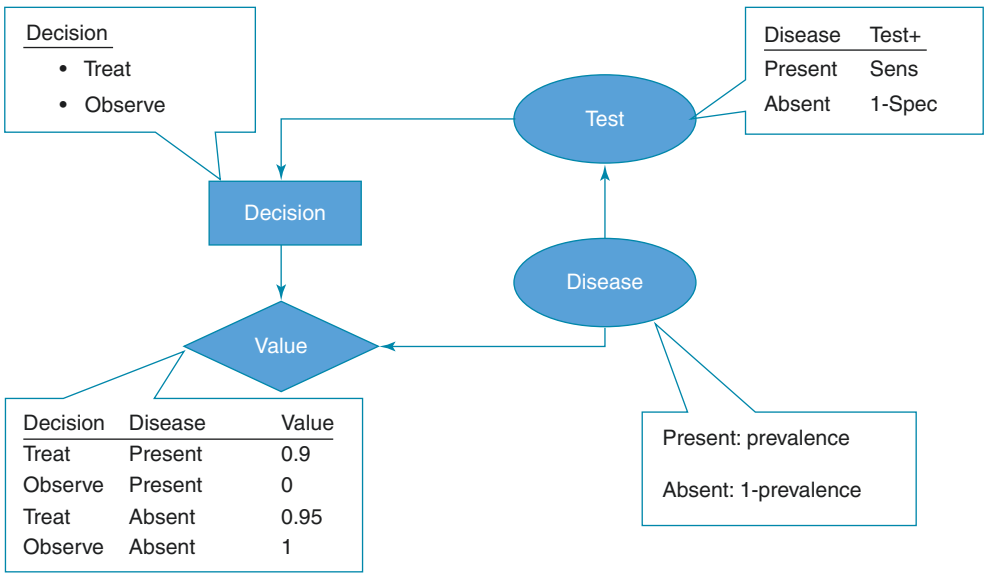

Fig. 6.10 A simple influence diagram. The diagram shows the three types of nodes found in an influence diagram: round *chance nodes*, a square *decision node*, and a diamond *value node*. The contents of each node are shown. An influence diagram with only chance nodes is known as a Bayesian belief network (or belief net)

influence diagram. It represents the decision to treat or observe given a test result and a prior probability of disease.

The round chance nodes represent random variables and store the probability distributions. The decision nodes store potential actions. The value node stores utilities for different possible states of the diagram. Arrows (also called arcs or edges) entering a decision node represent information available when the decision is made. In this case, the test result will be known before a treatment decision is made. Arcs going into a chance node represent variables on which the probabilities will be conditioned. The probability of a positive test result depends on whether the disease is present or not. Arcs going into the value node represent the variables that will affect the value of the diagram. In this model, the combination of the decision to treat or observe combined with the presence or absence of disease determines the value. The bubbles in Fig. 6.10 show the contents of each of the nodes.

Influence diagrams are useful for modeling complex relationships among random variables, often without decision or value nodes. An influence diagram composed of only chance nodes is also referred to as a Bayesian belief network (Bayes net or belief network). They are often used to make inferences on complex data, sometimes with hundreds of nodes. Inference engines that use Bayesian belief networks have been used to detect credit card fraud to complex diagnostic decision support [31, 32]. Bayesian networks in which the directed arcs have a strictly causal meaning are used in causal statistical analyses [33, 34].

One of the most recent applications of influence diagrams has been as a data structure for mobilizing computable biomedical knowledge to share decision support between sites in an executable format [35]. In this context, a decision model can represent a rule that determines what action to take under what circumstances. However, when the "rule" is represented as a full decision model, the various

parameters—probabilities and utilities—used to create the rule can be adjusted to local circumstances. So the "rule" can be tailored to the individual location.

Shared Decision Making

DA in the clinical setting was classically applied to a physician and patient facing a clinical decision where the outcomes are uncertain and high stakes [36]. However, it is rarely practical to complete a formal DA in the context of clinical care. It simply takes too much time and too many resources. For this reason, DA and CEA are generally used to guide practice in general or establish policy.

However, there is a clear need to address the components of a decision analysis at the bedside. This need has led to the emergence of **shared decision-making** (SDM) strategies [37]. SDM has become increasingly important as more clinical interventions emerge that are preference-sensitive [38]. The US Preventive Services Task Force has devoted a category of recommendation ("C recommendations") entirely for "selectively offering or providing this service to individual patients based on professional judgment and patient preferences" [39].

Most clinicians are familiar with shared decision-making in an informal sense, discussing the medical decision with a patient, providing an opportunity for the patient to ask questions and contribute to the decision. However, formal SDM is a more precise and nuanced process. SDM, sometimes called informed medical decision making, should meet three key requirements:

1. The patient is made aware of his or her options
2. The patient understands the likelihood of the important outcomes resulting from each option
3. The patient undergoes "values clarification," some exercise in which s/he expresses preferences over the outcomes

As a check, formal SDM may include a step in which the patient's final decision is evaluated as consistent or inconsistent with his or her expressed values. The elements of shared decision-making clearly correspond to the elements of DA, but the assessments and analyses are less quantitative.

SDM lends itself well to automation. There are large repositories of automated SDM tools, for example, at the Ottawa Hospital Research Institute [40], to which patients can be directed when they face one of these preference-sensitive decisions. Furthermore, there is a growing body of evidence that decision aids that automate SDM improve the quality of medical decision-making [41, 42]. Given the clear value-added from SDM, one might expect it to be incorporated broadly in EHRs [43]. However, to date, this has rarely been done [44].

Decision Support Prioritization

A more novel application of DA techniques to medical informatics and decision support is the prioritization of care recommendations for individual patients. This has been especially fruitful in the prioritization of preventive services. It was long ago well established that the number of preventive services recommended by authoritative bodies exceeds what can be done in a typical visit [45, 46]. Moreover, physicians are likely to spend precious clinical time on services with less value [47].

One strategy proposed to address this problem is to use decision-analytic algorithms to determine which preventive services offer the greatest expected value for the patient and prioritize decision support based on that calculation. Such a calculation would consider the likelihood the patient needs the relevant issue (prior probability), the seriousness of the issue (disutility), and the effectiveness of providing decision support to address it [48]. This approach has been demonstrated in both pediatric and adult settings [49–51]. By prioritizing decision support based on expected value, this approach can reduce alert fatigue while providing the most important decision support.

The Role of Decision Sciences in Clinical Informatics

Medicine is an information-intensive business rife with uncertainty, and humans are flawed data processors and decision-makers vulnerable to bias. Because computers can flawlessly and tirelessly process vast amounts of data, they have the potential, if used correctly, to compensate for these human frailties. But computers are only as correct as their programming. So a strong theoretical grounding for decision-making and decision support is indispensable.

Well-designed and well-executed decision models can form the basis of strong guidelines that you will want to be encoded in your systems. Models of complex Bayesian inference can help guide computer-based clinical decision support or represent decision rules in a format that can be readily adapted to new settings. DA approaches can also prioritize which decision support is provided, avoiding alert fatigue. Even day-to-day decision-making about IT purchases, investments, and distributions can be informed by more careful analysis of decisions made under uncertainty.

Future Directions

The relationship between decision analysis, guideline development, decision support, and quality measurement is growing continuously closer. There is a growing emphasis on

using EHRs and decision support to improve guideline adherence and measure the quality of care through quality indicators. Formal decision sciences techniques can improve every step in these processes.

Chapter Summary

In the clinical setting, you will often face difficult challenges that do not present clear, singular solutions. Maybe the 34-year-old woman has strep throat, or maybe she has seasonal allergies, or something much less common. Knowing the probability of each of these options is vital to effective treatment. Decision trees, expected utility theory, and other DA tools will help guide your decision-making process when deciding on the best course of action for each of your patients. More and more, these theories and models are adopted by technology to create computerized clinical decision support systems. Because a computer can process much more information at a much faster speed than one physician, CDSS can be invaluable in providing an efficient and effective medical practice.

Questions for Discussion

1. A healthy 56-year-old patient presents with influenza-like illness (ILI). How could you apply Bayes Theorem to update the probability of the patient having COVID-19 versus another ILI as you gather evidence?
2. Under what circumstances could a computer make a reliable diagnosis by applying Bayesian algorithms? Would a system be acceptable to patients? Physicians? Payors?
3. Think about a time in the past when your own potential biases or heuristics in probability estimation influenced your decision-making. Did they help you make a correct diagnosis or prevent you from making the right diagnosis?
4. A new CMO at the medical center suggested that primary care clinicians should begin applying formal decision analysis when treating patients to enhance shared decision-making (SDM). As the CMIO, would you support this recommendation? Why or why not?
5. A hospital board member pulls you to the side after your presentation on a new CDS system that uses Markov chains. He says that he believes Markov chains are based on utilitarianism, which he views as un-American. He also expresses concerns about potential Russian influence on the hospital's information system. How might you politely set him straight about the use of Markov models in the CDS system to support rationale decision-making by clinicians?

References

1. von Neumann J, Morgenstern O. Theory of games and economic behavior. Princeton, NJ: Princeton University Press; 1944.
2. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. Econometrica. 1979;47(2):263–92.
3. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. Science. 1974;185(4157):1124–31.
4. Howard RA. Decision analysis: applied decision theory. In: Howard RA, Matheson JE, editors. Readings on the principles and applications of decision analysis. Menlo Park, CA: Strategic Decisions; 1983. p. 95–113.
5. Bernstein PL. Against the Gods: The remarkable story of risk. New York: Wiley; 1998.
6. Kolmogorov AN. Foundations of the theory of probability. Second English ed. New York: Chelsea; 1956.
7. Hajek A. Interpretations of probability. In: The Stanford encyclopedia of philosophy. 2012. https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/.
8. Venn J. The logic of chance. London: MacMillan; 1876.
9. Centor RM, Witherspoon JM, Dalton HP, Brody CE, Link K. The diagnosis of strep throat in adults in the emergency room. Med Decis Making. 1981;1(3):239–46.
10. McIsaac WJ, Kellner JD, Aufricht P, Vanjaka A, Low DE. Empirical validation of guidelines for the management of pharyngitis in children and adults. JAMA. 2004;291(13):1587–95.
11. American Heart Association. Go red for women. Common myths about heart disease. 2021. https://www.goredforwomen.org/en/about-heart-disease-in-women/facts/common-myths-about-heart-disease.
12. Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Communicating statistical information. American Association for the Advancement of Science; 2000.
13. Joyce J. Bayes' theorem. In: The Stanford encyclopedia of philosophy. 2003. http://plato.stanford.edu/archives/fall2008/entries/bayes-theorem/.
14. Downs SM, Walker MG, Blum RL, editors. Automated summarization of on-line medical records. In: MEDINFO '86, 1986, North Holland, Amsterdam.
15. Warner HR, Haug P, Bouhaddou O, Lincoln M, Warner Jr H, Sorenson D, et al., editors. ILIAD as an expert consultant to teach differential diagnosis. In: Proceedings of the annual symposium on computer application in medical care. American Medical Informatics Association; 1988.
16. Miller RA. Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. J Am Med Inform Assoc. 1994;1(1):8–27.
17. Downs SM. Interval likelihood ratios and the diagnostic process. Pediatrics. 2021;147(1). https://doi.org/10.1542/peds.2020-037721.
18. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med. 1980;302(20):1109–17.
19. Bernoulli D. Exposition of a new theory on the measurement. Econometrica. 1954;22(1):23–36.
20. Raiffa H. Decision analysis: introductory lectures on choices under uncertainty. New York: Random House; 1968.
21. Howard RA. Risk preference. In: Howard RA, Matheson JE, editors. Readings on the principles and applications of decision analysis. Menlo Park, CA: Strategic Decisions; 1983.
22. Torrance GW. Utility approach to measuring health-related quality of life. J Chronic Dis. 1987;40(6):593–600.
23. Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. Oper Res. 1980;28(1):206–24.
24. Weinstein MC, Russell LB, Gold MR, Siegel JE. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.

25. Torrance GW, Feeny D. Utilities and quality-adjusted life years. Int J Technol Assess Health Care. 1989;5(4):559–75.
26. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI®): concepts, measurement properties and applications. Health Qual Life outcomes. 2003;1(1):1–13.
27. Coca Perraillon M, Shih Y-CT, Thisted RA. Predicting the EQ-5D-3L preference index from the SF-12 health survey in a national US sample: a finite mixture approach. Med Decis Making. 2015;35(7):888–901.
28. Chapman RH, Stone PW, Sandberg EA, Bell C, Neumann PJ. A comprehensive league table of cost-utility ratios and a sub-table of "panel-worthy" studies. Med Decis Making. 2000;20(4):451–8.
29. Nease RF Jr, Owens DK. Use of influence diagrams to structure medical decisions. Med Decis Making. 1997;17(3):263–75. https://doi.org/10.1177/0272989x9701700302.
30. Owens DK, Shachter RD, Nease RF Jr. Representation and analysis of medical decision problems with influence diagrams. Med Decis Making. 1997;17(3):241–62. https://doi.org/10.1177/0272989x9701700301.
31. Dheepa V, Dhanapal R. Analysis of credit card fraud detection methods. Int J Recent Trends Eng. 2009;2(3):126.
32. Shwe MA, Middleton B, Heckerman DE, Henrion M, Horvitz EJ, Lehmann HP, et al. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base. Methods Inf Med. 1991;30(04):241–55.
33. Ankan A, Wortel IMN, Textor J. Testing graphical causal models using the R package "dagitty". Curr Protoc. 2021;1(2):e45. https://doi.org/10.1002/cpz1.45.
34. Pearl J, Glymour M, Jewell NP. Causal inference in statistics: a primer. Sussex, UK: Wiley; 2016.
35. Lehmann HP, Downs SM. Desiderata for sharable computable biomedical knowledge for learning health systems. Learn Health Syst. 2018;2(4):e10065. https://doi.org/10.1002/lrh2.10065.
36. Pauker SG. Clinical decision-making rounds at the Tufts—New England Medical Center Hospital. Med Decis Making. 2005;25(1):81–2.
37. Kuppermann M, Sawaya GF. Shared decision-making: easy to evoke, challenging to implement. JAMA Intern Med. 2015;175(2):167–8.
38. Wennberg JE. Tracking medicine: a researcher's quest to understand health care. New York: Oxford University Press; 2010.
39. U.S. Preventive Services Task Force. Grade definitions. 2018. https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes/grade-definitions. Accessed Sept 2021.
40. The Ottawa Hospital Research Institute. Patient Decision Aids. 2020. https://decisionaid.ohri.ca/. Accessed 13 Sept 2021.
41. Stacey D, Légaré F, Col NF, Bennett CL, Barry MJ, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2014;(1):Cd001431. https://doi.org/10.1002/14651858.CD001431.pub4.
42. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2017;4(4):Cd001431. https://doi.org/10.1002/14651858.CD001431.pub5.
43. Day FC, Pourhomayoun M, Keeves D, Lees AF, Sarrafzadeh M, Bell D, et al. Feasibility study of an EHR-integrated mobile shared decision making application. Int J Med Inform. 2019;124:24–30. https://doi.org/10.1016/j.ijmedinf.2019.01.008.
44. Kuo AM, Thavalathil B, Elwyn G, Nemeth Z, Dang S. The promise of electronic health records to promote shared decision making: a narrative review and a look ahead. Med Decis Making. 2018;38(8):1040–5. https://doi.org/10.1177/0272989x18796223.
45. Belamarich PF, Gandica R, Stein RE, Racine AD. Drowning in a sea of advice: pediatricians and American Academy of Pediatrics policy statements. Pediatrics. 2006;118(4):e964–e78.
46. Yarnall KS, Pollak KI, Østbye T, Krause KM, Michener JL. Primary care: is there enough time for prevention? Am J Public Health. 2003;93(4):635–41.
47. Pollak KI, Krause KM, Yarnall KS, Gradison M, Michener JL, Østbye T. Estimated time spent on preventive services by primary care physicians. BMC Health Serv Res. 2008;8(1):1–7.
48. Downs SM, Uner H. Expected value prioritization of prompts and reminders. Proc AMIA Symp. 2002:215–9.
49. Anand V, Carroll AE, Biondich PG, Dugan TM, Downs SM. Pediatric decision support using adapted Arden Syntax. Artif Intell Med. 2018;92:15–23.
50. Downs SM, Bauer NS, Saha C, Ofner S, Carroll AE. Effect of a computer-based decision support intervention on autism spectrum disorder screening in pediatric primary care clinics: a cluster randomized clinical trial. JAMA Netw Open. 2019;2(12):e1917676-e.
51. Taksler GB, Keshner M, Fagerlin A, Hajizadeh N, Braithwaite RS. Personalized estimates of benefit from preventive care guidelines: a proof of concept. Ann Intern Med. 2013;159(3):161–8.