<1> Chapter 2: Cost-benefit and cost-effectiveness analysis

In this Chapter, we seek to answer the question: how much should we pay for a public health program? We often have to decide how to allocate funds to different public health programs, or decide whether a new medical test or treatment is worth the cost. How can we make such decisions fairly? We'll first work through some examples of *decision trees* that are commonly used to make these judgments in a rigorous and fair way. We'll create some decision tress to perform *cost-benefit analysis* to determine whether we want to pay for a new service, test, or treatment if we are focused on lowering the costs of our operations. We'll then understand how to perform *cost-effectiveness analysis* to identify under what circumstances a more expensive new service, test, or treatment might be worth the cost because it meaningfully improves health outcomes.

<2>Decision trees

We often have to make complex decisions about what course of action to pursue; in public health and healthcare service delivery, those decisions can have life-or-death consequences. To make the best decisions based on the information we have available, we often construct *decision trees*—which are graphical representations of a decision and its potential consequences. Decision trees represent one of the most basic strategies for public health and healthcare problem solving, and serve as a foundation for the future models that we will develop to analyze more complex decisions.

To understand how to build and solve a decision tree, suppose we have a common decision problem: whether to administer a new, experimental drug to a patient. Suppose a new, experimental drug has been designed to treat a deadly form of kidney cancer. If a person undergoes treatment for the cancer with a standard, existing drug, they have a 15% chance of survival and an 85% chance of death. With the new, experimental drug, they have a 40% chance of remission (tumor shrinkage to an undetectable level) and a 60% chance of death. However, among those people who experienced a remission, 50% survived and 50% later died due to cancer recurrence.

If we wish to determine whether to recommend the new, experimental drug to a patient with kidney cancer, how might we rationally compare the standard, existing drug to the new, experimental drug?

To solve this problem, we can draw all possible outcomes on a *decision tree*, which specifies both the probability of each outcome and the "value" of each outcome (see Figure 2.1). For this problem, we value the outcome of survival, so we will specify that an outcome has value 0 if a person dies and value 1 if a person survives; this designation will allow us to easily calculate and compare the probability of survival between the two groups.

[INSERT FIGURE 2.1 HERE]

As shown in Figure 2.1, we typically designate the start of a decision tree with a square, and each "chance node" or decision branching point with a circle. Triangles then designate the endpoint (sometimes called a leaf node or terminal node) of each branch. Next to each decision

branch is the probability of traveling along that branch. Next to each node is the value of the node.

To "solve" the decision tree, we start from the right-hand side of the tree and calculate the probability of survival for each node, "rolling back" ultimately to the first node (the root node) to determine the ultimate expected probability of survival for our patient if they were offered one therapy or another.

For example, suppose we start with the top node corresponding to the standard, existing drug. We see there is an 85% probability of death and a 15% probability of survival, leading to a "value" (probability of survival) of 0.85 x 0 + 0.15 x 1 = 0.15.

Similarly, if we start with the bottom-right node we can see that with remission that there is a probability of survival of 0.5. Because there is a 60% probability of death and a 40% probability of remission with the new, experimental drug, we can calculate that the overall probability of survival with the new, experimental drug is 0.6 x 0 + 0.4 x 0.5 x 1 = 0.20 (the 0.4 x 0.5 is the probability of remission multiplied by the probability of survival given remission), as shown in Figure 2.2.

[INSERT FIGURE 2.2 HERE]

The "solved" decision tree makes our decision fairly clear: we should recommend the new, experimental drug to our patient, assuming that survival is the primary outcome we are basing our decision on.

<2>Cost-benefit analysis

Solving a decision tree is a nice way to organize our thinking, and often helps us make decisions that are more complex than calculating the probability of survival. Many decisions in public health and healthcare settings involve complex decisions about whether to undertake a potentially-expensive new course of action. In these contexts, decision trees can help us organize the probability of experiencing a budget deficit or budget excess following a critical organizational decision.

Suppose you direct a medical clinic that provides outpatient primary healthcare services. Most patients in the clinic see a physician for their visits, after which the patient's insurance company is charged a fee. The fee is based on whether the visit was of low, medium, or high complexity, which corresponds to the complexity of diseases the patient is being treated for and the amount of time required to conduct the visit.

The clinic is strapped for cash and can't afford to experience any reductions in revenue at the moment. But you need to make a critical financial decision: whether or not to participate in a new clinical program. The new program offered by insurance companies involves changing the way low complexity patients are managed. Under the new program, insurance companies would pay the clinic to have fewer in-person visits with low complexity patients; instead of having the clinic doctors visit these patients in person, the doctors could confer with these patients through a computer screen, having a 'virtual visit' (paying $10 per visit). The low complexity in-person visits replaced by virtual visits would specifically focus on run-of-the-mill colds or heartburn symptoms, not life-threatening conditions. The virtual visits would only take about half as long

as in-person visits, so your clinic doctors can see more patients of moderate and high complexity, while spending less time on patients with low complexity problems.

Should you sign up your clinic for the new program?

Here's the data you have available:

- A quick poll indicates that one-third of low complexity physician visits would be replaced by a virtual visit, while the other two-thirds of low complexity physician visits would continue to take place as in-person visits, given patients' preferences.
- Under the existing system, about 30% of visits for each physician in your clinic are of low complexity (paying $50 per visit), 50% are of medium complexity ($100 per visit), and 20% are of high complexity ($150 per visit).
- Under the new virtual visit program, instead of 30% of in-person physician visit time being taken up by low complexity visits, only 20% of physician time (two-thirds) will be taken up by low complexity in-person visits, and the other low complexity visits will be half as long, taking a total of 5% of time.
- Hence, the virtual visit program will leave 5% of physician time to be split between more moderate and high complexity patient visits.

Even though virtual visits have a lower payment rate (paying $10 per visit) than the in-person low-complexity visit (paying $50 per visit), will the clinic stand to gain or lose money by converting to virtual visits for low complexity patients? In other words, will the increase in moderate and high complexity visits outweigh the revenue loss from low complexity in-person visits being converted to virtual visits?

We can draw a decision tree to help answer these questions. As shown in Figure 2.3, the status quo scenario can be drawn as the proportion of visits for each of low, moderate, and high complexity visits, multiplied by the typical payment rate for each of these, before the virtual visit program. Seeing the top part of the decision tree in Figure 2.3, we see that the status quo scenario makes 0.3 x $50 + 0.5 x $100 + 0.2 x $150 = $95 for the clinic on average per visit.

[INSERT FIGURE 2.3 HERE]

As shown in the lower branch of Figure 2.3, the scenario of converting one-third of low complexity in-person visits to virtual visits is more complex. The conversion has two effects: first, one-third of low complexity visits are removed and now take up only 5% of the physician's time and are paid only $10 instead of $50; second, the 5% additional time is distributed evenly between moderate and high complexity visits (which increase by 2.5% each in time use). Hence, the new scenario makes 0.2 x $50 + 0.05 x $10 + 0.525 x $100 + 0.225 x $150 = $96.75 for the clinic on average per visit.

So, if financial viability was the only consideration, the virtual visit program would be a wise decision, as it would provide the clinic an additional $1.75 per visit, on average.

<2> Cost-effectiveness analysis

In the previous section, we performed *cost-benefit analysis*, which is the process of synthesizing available information to determine which decision we should make to minimize our overall financial losses in providing a health service.

In many public health and healthcare decisions, however, our goal is not to minimize overall losses, but to spend funds wisely even if we incur a financial cost. Even if we spend more money, the benefits in terms of reducing morbidity (cases of disease) and mortality (deaths from disease) may be justified. Typically, we use *cost-effectiveness analysis* to identify how much we reduce disease morbidity or mortality, per dollar spent; in contrast to cost-benefit analyses, cost-effectiveness analyses are not just trying to find the strategy with the lowest overall cost, but rather trying to find the strategy that maximizes how much we improve health per each dollar spent.

To illustrate the distinction between these two forms of analysis, suppose we took the example of studying a new therapy for tuberculosis. The standard treatment for tuberculosis in most of the world is prolonged, lasting at least six months, and often carries significant side-effects for patients. Suppose it costs about $10,000 total. Also suppose a new drug is available that we can add to the standard treatment for tuberculosis. The new drug makes the overall treatment regimen just as effective, but slightly more costly (about $2,000 more for the overall treatment, or $12,000 total). The combined treatment regimen will have more side effects because of the additional drug, but will also be shorter, just four months. How do we make a fair comparison and evaluate whether this new treatment is "worth" the cost? To put it bluntly: is paying $2,000 more for the treatment "worth" having two months shorter duration of treatment?

The first challenge we face is that the question is no longer purely in terms of dollar amounts; we now have to quantify human suffering, which is an entirely different, and more challenging, enterprise. How do we measure how much less suffering our patients will endure if they have to face side-effects for only four months instead of for six months?

One long-standing measure for human suffering is the quality-adjusted life-year (QALY, pronounced "qual-ee", as in "quality"), which was devised in the 1960's to create a fair measure of how much suffering people endured among different diseases or treatments. The concept of a QALY is illustrated in Figure 2.5. Suppose that we have a perfectly healthy person, and that person lives for one year in a state of perfect health; we say that person has gained 1 QALY. Suppose a person is dead in a given calendar year; that person has gained 0 QALYs. In between the two are most people with disease, who either are not perfectly-healthy (their quality of life, known as their *utility* or *utility weight*, is not a perfect score of 1) or they don't survive for the entire year (the duration of time that they are alive and able to accumulate utility is less than 1 year).

A QALY is therefore a product of two entities: the utility or quality of a person's life in a year, and how many years they are in that utility state/at that quality level. Suppose a person is suffering badly from a disease or from side-effects of a treatment for a short period of time. They may have a low utility (like 0.2 rather than the perfectly-healthy utility value of 1), and have that disease for a short period such as 6 months (0.5 of a year), then go back to being perfectly healthy for the rest of the year. Their overall QALYs gained over the year might therefore be 0.2 utility x 0.5 years + 1.0 utility x 0.5 years = 0.6 QALYs gained over the course of the entire year. Suppose, however, that another person has a less-severe disease but that it lasts longer; for example, they may have a moderate utility (like 0.6) but have the disease for the entire year. Their overall QALYs gained over the year would therefore be 0.6 utility x 1 year = 0.6 QALYs. As shown in this example, there is an important caveat to know about QALYs: the overall QALYs accumulated don't provide us with clear information about the disease or treatment. Rather, someone with a brief but severe disease is counted the same way as someone with a less

severe but prolonged disease (note, later, that we talk about “discounting” as a strategy to partly address this issue, later in this chapter).

Where do utility weight values come from? There are at least four common methods that scientists use to estimate the utility weights for diseases or treatments. All of these methods utilize surveys of patients with the disease, or surveys among people who are healthy but presented with a situation in which they have to make choices about how they would handle a future disease. One type of experiment is known as the “standard gamble”, which presents survey respondents with the task of choosing between having a disease for a longer period of time, or having a treatment that would either provide them with perfect health or kill them. The worse the disease (the lower the utility weight), the more people are willing to risk undergoing the potentially-deadly treatment for it, rather than staying unhealthy. The standard gamble thus allows scientists to estimate how bad different diseases are, relative to each other, ranking them from 0 to 1 in terms of utility weights. Another measurement approach is called the “visual analogue scale”, which involves asking people with the disease or undergoing a particular treatment with side-effects to rate their health from 0 to 100, then divide the score by 100 to get a utility weight between 0 and 1. The third measurement approach is called a “time trade-off” study, in which people are asked to theoretically choose between being ill with a given disease (or being on a given treatment for a period of time), or having perfect health for a shorter period but having shorter life-expectancy. As you can imagine, such theoretical experiments are often difficult to believe, as people actually experiencing a disease would not be expected to react in the same way as someone imagining a theoretical illness. Furthermore, a person’s willingness to undergo various treatments may change over time and with age, as well as from day to day or month to month. Hence, estimating utility values is fundamentally difficult and controversial;

important debates have taken place about how to better quantify people's state of health. The fourth and increasingly common strategy to measure utility is to administer a questionnaire that asks, in a standardized way, about many different domains of life: the ability to participate in daily life activities (such as work and leisure activities), care for oneself (e.g., go to the bathroom or dress oneself without assistance), and be mobile (e.g., walk without pain or shortness of breath). The questionnaire also assessed a person's degree of discomfort or pain, as well as degree of anxiety or depression. By standardizing such measures, utility values can be calculated similarly across many different diseases or conditions, without resorting to theoretical trade-offs.

Getting back to our tuberculosis example, we can imagine that we can plot out the utility of a person's life while undergoing tuberculosis treatment, versus how long they have to undergo treatment, as shown in Figure 2.4.

[INSERT FIGURE 2.4 HERE]

As shown in Figure 2.4, the utility measure can help us quantify the two key differences between our standard tuberculosis treatment and the new treatment. The standard treatment lasts for six months and puts people into a moderate state of quality of life (0.8) because of its side-effects. The newer drug that modifies the treatment regimen makes the overall treatment course shorter (totaling four months) but puts people into a lower state of quality of life (0.75) during that period, because of more severe side-effects.

From Figure 2.4, we can calculate the expected total number of QALYs under the standard treatment, then under the new treatment.

Under the standard treatment, over one year a patient would gain 0.8 utility x 0.5 year + 1.0 utility x 0.5 year = 0.90 QALYs. Under the new treatment, over one year a patient would gain 0.75 utility x 0.333 year + 1.0 utility x 0.666 year = 0.92 QALYs.

Is this gain in QALYs with the new treatment "worth it"? A typical way of quantifying the value of QALYs gained is to specify the *incremental cost-effectiveness ratio (ICER)*, which is defined in Equation 2.4.

[Equation 2.4] $$ICER = \frac{(New\ Cost - Old\ Cost)}{(New\ QALYs - Old\ QALYs)}$$

In this case, the ICER is ($12,000 - $10,000) / (0.92 – 0.90) = $100,000 per QALY gained. Is this a good value for money? In prior years, it was common to define that an intervention would be considered "cost-effective" if it cost less than $50,000 per QALY gained. But this threshold has been recognized as inherently arbitrary, particularly because QALYs themselves are so difficult to measure. In more recent years, scientists have tried to identify what different governments or communities are "willing to pay", based on what is actually paid for. Others have argued that a willingness to pay threshold should be based on the typical gross domestic product (GDP) of a country, which would potentially discriminate against people in poverty.

Rather than setting an absolute threshold to declare an intervention "cost effective" or "not cost effective", many scientists will now use ICERs to simply compare different interventions and clearly specify whether a treatment is more or less effective, and more or less costly, than an alternative, to better inform policy-makers. A common way to differentiate alternative treatments is to plot results on a cost-effectiveness plane, shown in Figure 2.5.

[INSERT FIGURE 2.5 HERE]

The plane illustrates key terms that we commonly use to define how we interpret the cost-effectiveness of a new treatment compared to an existing one. The x-axis quantifies how effective a treatment is, in QALYs. The y-axis quantifies how costly the treatment is, in dollars. The upper right quadrant is the most common quadrant where we find new treatments: it identifies treatments that are more effective but also more costly. The diagonal line cutting through this quadrant identifies the willingness to pay threshold—or how much people are willing to pay for each additional QALY (e.g., $50,000 per QALY, classically). The upper left quadrant is less effective but more costly—a situation in which we say that the standard treatment *dominates* the new treatment, hence the new treatment is *rejected* or *excluded* (not desirable). The bottom right quadrant is more effective and less costly—a fantastic but rare circumstances in which we say that the new treatment *dominates* the standard treatment, hence the new treatment should be widely accepted. Finally, the bottom left quadrant indicates when a treatment is less effective but also less costly. Most of the time such a treatment would be questionable at best, as we wouldn't want to save money at the expense of causing more morbidity or mortality.

<2> Common terms and metrics

Sometimes scientists will use the term *cost-utility analysis* to refer to the type of analysis we just performed, in which we calculated an ICER in terms of dollars spent per QALY gained. The concept is meant to highlight that we have incorporated utility values into our analysis. These scientists might refer to cost-effectiveness analysis as a situation in which we do not adjust

the life-years gained with utility values, but simply analyze total life-years gained from one treatment versus another. The problem with such a cost-effectiveness analysis is that it only looks at gains in life expectancy, not in whether that life expectancy gain is painful and full of suffering, or healthy and happy. Often the term cost-effectiveness analysis is applied widely without clarifying that most cost-effectiveness analyses in the literature are actually cost-utility analyses, because they consider QALYs rather than absolute numbers of years of life gained.

Similarly, other critics will choose to tabulate the utility value of a treatment in terms of disability-adjusted life-years (DALYs, pronounced "dah-lees") rather than quality-adjusted life years. DALYs are conceptually different from QALYs because they focus on a loss rather than a gain: DALYs focus on how much life is lost due to disability and early death from a disease. Therefore, a good intervention will reduce the number of DALYs lost to a disease. While QALYs are calculated as a utility value times the number of years lived at that utility value, DALYs are calculated as the sum of two components as shown in Equation 2.5.

[Equation 2.5] DALYs = Years of Life Lost + Years of Life Lived with Disability

Years of life lost are calculated per Equation 2.6.

[Equation 2.6] Years of Life Lost = Number of deaths due to condition x (Life expectancy – Age of death from condition)

Years of life lived with disability are calculated per Equation 2.7.

[Equation 2.7] Years of life lived with disability = Incidence of disease in population x Disability weight for the disease x Duration of the disease until remission or death

In older tabulations, DALY's were often calculated by including different disability (disutility) weights for different age groups, such that losses of life from older adults were not counted as heavily as those among younger adults. This practice became controversial and is no longer commonly performed.

DALYs are currently used by the World Health Organization to help tally the burden of disease in an overall population, and therefore how much an intervention reduces the burden of disease in an overall population rather than just for a single individual. The practice brings up the issue of *perspective* in cost-effectiveness analyses, which refers to the idea that for any given analysis in which we are tabulating costs and/or effectiveness (using any measure, such as QALYs or DALYs), we have to define who we are doing the analysis *for*. For example, an individual person in the United Kingdom may not pay many of the costs of their medical treatment, because those costs are covered by the UK National Health Service, typically (with the caveat that the person indirectly pays by paying taxes to the government, who then pools the funds). But most cost-effectiveness analyses are performed from the "societal" perspective, which means that—regardless of who pays—we wish to know as a society how much is being gained in paying for a new test, intervention, or treatment, and how much is being lost. Some analyses may nevertheless be made from other perspectives, such as the perspective of an individual government agency or an insurance company.

Finally, a common practice in cost-effectiveness analysis is to "discount" both the costs and QALY/DALY estimates tabulated in a cost-effectiveness analysis. *Discounting* or *temporal discounting* refers to the finding in behavioral economics that people tend to value money and health in the short-term more than they value money and health in the long-term. For example, most people choose to receive $100 immediately rather than receiving $100 one month from

now, even though both choices are worth the same amount of money. To consider the fact that people tend to favor immediate money savings or immediate health gains more than distant ones, cost-effectiveness analyses typically discount each year of costs or QALYs considered in an analysis by a small factor (usually around 3% by convention) per Equation 2.8.

[Equation 2.8] Current value = Future value/[(1+discount rate)^(years in the future)]

For example, if applying a 3% annual discount rate to a calculation of costs, a therapy that costs $100 next year will be calculated as costing $100/(1.03^1) = $97, as compared to a therapy that costs $100 this year, which will remain valued at $100 in present-day dollars. Similarly, a therapy that prevents disease five years in the future and gains 0.1 QALYs in five years would be worth 0.1/(1.03^5) = 0.09 QALYs in present-day terms, as compared to a therapy that saves 0.1 QALYs this year and remains valued at 0.1 QALYs. Typically, cost-effectiveness studies tend to calculate and sum all costs and QALYs over a long-term “time horizon”, usually the overall life expectancy of a person, to compare all costs and health states in the future. But discounting has been controversial, because it may undervalue the benefits of preventive interventions that avert long-term disease, and favor short-term therapies instead.