**UNIT 7: Evaluating Care**

Overview:

7A:. The video reviews Incremental Cost Effectiveness Ratio (ICER) calculations, noting the need for a broad definition of costs and benefits.

7B:. The video reviews the logic for and options for calculation of Quality Adjusted Life Years as an input to evaluation of treatment options. Difficulties with affective forecasting complicate the measurement and interpretation of QALYs.

Learning Objectives:

1. Understand how to calculate ICERs to compare treatments in terms of cost effectiveness
2. Understand QALY calculations, including the benefit of scaling both longevity and health-related life quality into one metric but also limitations due to difficulties with affective forecasting.

Role in Course: The content in this section is intended as a reminder that we must look carefully at potential treatment options to determine value. Across the course, any mentions of striving for or rewarding “more valuable care” rest on an underlying assumption that rigorous cost-effectiveness analysis exist to tell us where that value lies. Note that you saw a bit of this approach in David’s Bootcamp course. Further, Ryan’s Analysis of Healthcare Effectiveness and Outcomes course provides sophisticated tools for assessing treatment value. (Unit 8 of this course also complements Ryan’s course.)

**Introduction: Evaluating Treatment Options**

Cost Effectiveness Analyses (CEA) often influences advice and guidelines to providers. Even in situations where cost is not focal, careful CEA rests on understanding incremental benefits, and this is important for ensuring quality of care. Further, payers require good CEA information to best set incentives. Finally, patients can benefit from CEA-driven advice that supports individual health-related decision making and behavior.

While we can use CEA to evaluate any care, the techniques are often focused on care that we believe is high benefit but is also high cost (e.g., many new technologies). I’ll focus on Incremental Cost Effectiveness Ratios or ICERs as a useful methodology for conducting CEA, and then I will review calculations of Quality Adjusted Life Years, or QALYs, as an input to ICER analyses.

**Marginal Analysis**

CEA rests on a basic insight from microeconomics: if we want to make sound decisions, we need to attend to the incremental costs and benefits of activities. This is a way of thinking that makes rational sense but can sometimes be forgotten in health policy discussions. We sometimes get distracted by huge absolute numbers rather than thinking more carefully, and incrementally, about the marginal impacts of the real decisions that are currently under consideration. For instance, Cutler and McClellan *Health Affairs* 2001 use CEA to illustrate that the $5 billion US Spending on heart attacks in 1998 (up from $3 billion in 1984) returned, on average, one year of additional life for each $10,000 in additional spending. While the absolute increase in spending seems distressing ($2billion more!), marginal calculations seem much more reasonable ($10,000 per additional life year). More specifically, CEA simply involves considering the costs versus the outcomes (or effects) of courses of action; calculating ICERs is a major way to conduct CEA.

**ICER**

For mutually exclusive clinical treatments, we can use incremental cost effectiveness analysis or ICER to compare options. The formula is simply:

(C1-C2) / (B1-B2)

Where:

* C1 is the net cost of a new treatment under consideration
* B1 is the net benefit of that new treatment.
* C2 and B2 refer to net costs and net benefits of a reference treatment (often the current standard of care, but potentially a different treatment or even a scenario of no treatment)

For example, imagine a disease that causes a quick, painless, earlier death, with treatment options, costs, and benefits described in the table.

|  |  |  |
| --- | --- | --- |
|  | **Life Expectancy Given Treatment Option** | **Intervention Cost** |
| **Option 1: Reference Scenario of no Treatment** | 70 years | $0 |
| **Option 2: Drug Therapy** | 75 years | $5,000 |
| **Option 3: Surgery** | 75.25 years | $30,000 |

Imagine there aren’t material differences in suffering or disability, the only relevant benefit to consider is life expectancy. Imagine patients can have the drug or surgery, but not both.

Given our scenario, there are three comparisons we could make using ICER:

|  |  |  |
| --- | --- | --- |
|  | **No treatment** | **Drug** |
| **Drug** | ($5k-$0) / (75yr – 70yr) =  $1,000 per year | N/A |
| **Surgery** | ($30k-$0) / (75.25yr – 70yr) =  $5,714 per year | (30k-$5) / (75.25yr – 75yr) =  $100,000 per year |

So, for instance, the ICER in the upper-left quadrant tells us that Drug treatment, as compared to No treatment, costs $1,000 per year of life saved.

**Structuring ICER Problems**

ICER is only useful if it includes all important costs and benefits that differ between the two alternatives under consideration. Any costs (or benefits) that are the same for the two alternatives being considered in ICER will just subtract out to zero. What is key is to isolate unique costs and benefits.

It’s often important to have input from providers and patients who will have a comprehensive sense of all of the relevant costs and benefits. For this reason, with major ICER initiatives, there will often be a few iterations where patients, providers, and other stakeholders are allowed to comment on treatment experiences, disease burden, etc.

ICER usually assumes **all relevant costs** are included. Typically, we consider costs from a broad, societal perspective. Some examples might include the following:

* The costs of delivering care is usually the most obvious and important cost. This includes fees for interventions. It also includes the entire time course of treatment costs, such as provider time for diagnosis and follow up, if these are unique to one alternative. Many treatment options directly require future follow up, often over long time periods, so these should be considered as well.
* The costs to the patient, such as transportation costs, the average value of patient time spent in treatment, and the opportunity cost of lost productivity are also relevant. Often these opportunity costs are estimated based on population characteristics such as the average wage.
* One final point to note about costs is that good decision making requires adjustment for the time value of money (i.e., discounting future cash outlays).

In terms of benefits, again ICER assumes a broad approach. Benefits are often more challenging than costs, however, because they are more variable and hence more complicated to monetize or put into any particular units.

* The most obvious and important benefits are usually clinical patient outcomes in terms of longer life expectancy and disability avoided. (Far off benefits could be discounted if desired.)
* We might also consider more auxiliary benefits as well. For instance, being declared cured of cancer might decrease ongoing anxiety. If this is a pervasive benefit experienced by patients, one might choose to include the value of that benefit as well. Of course, the wider the range of benefits we consider, the harder they usually become to estimate and value.

**QALY Calculations**

To maximize the value of healthcare, we need to know the benefits of various treatment options. More precisely, if we want to conduct ICER, we need to scale treatment benefits into general, numerical terms. QALY calculations are a method for doing so. The main advantage of QALYs is that they put both disability and life expectancy into one common metric – a metric we can use in the denominator of ICER ratios.

An underlying assumption of QALY measurement is that we can arrange summary health-related states of being on a scale. The numbers assigned to QALY scales don’t really matter, as long as we are consistent across various calculations being compared. Typically we use a 0 (death) to 1 (perfect health) scale. One extra year of perfect health is equivalent to one QALY. An extra year at 50% health (or equivalently 50% disability) is equivalent to.5 QALY.

The difficulty with this method arises because we have to ask people to scale different expected health states into QALY terms. There are many different ways of asking these questions, and all have pros and cons.

One common QALY method is the **time-tradeoff measure**. For example, a researcher might ask might ask participants:

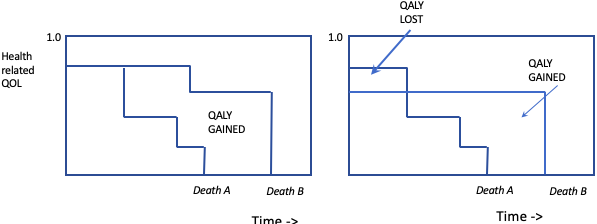
* Imagine that you have 10 years left to live and you will spend these 10 years lucid and in no pain but paralyzed from the waist down and dependent on a wheelchair.
* Imagine that you could choose to give up some life years to live for a shorter period in full health, specifically not wheelchair dependent.
* Indicate the number of years in full health that you think is of equal value to 10 years in the state described:

10 years wheelchair-dependent = \_\_**?**\_ years perfect health

Essentially, the participant is asked to scale disability into life year terms. For instance, if a participant said that 7 years at full health would be equivalent to 10 years in a wheelchair, this indicates that a year in a wheelchair is worth .7 QALYs (7/10 = .7).

Another way to think about this is the participant has been asked to set two outcomes equal to one another. The problem starts by comparing 10 years in a particular health state denoted by “X” QALYs (here, the health state of needing a wheelchair) to an unspecified number of years in full health. This is essentially the following equation: 10 years \* X QALY = \_\_?\_ years \* 1.0 QALY. If the participant answers 7, we have: 10 years \* X QALY = 7 years \* 1.0 QALY. This reduces to 10X = 7 or X = 7/10. So, the respondent is telling us that ten years in the state described is equivalent to 7 years of full health.

The graphs below illustrate the general logic of QALYs. On each graph, we see the time course of both health-related quality of life (QoL) and longevity. Each graph shows a comparison between Treatment A and Treatment B.



In the first graph, treatment B has two advantages: It extends life, and it also increases the quality of life. With treatment A, health related quality of life starts to degrade earlier, and drops twice before death. So, the region between the two lines represents the QALYs gained with treatment B.

In the second graph, the benefits of treatments A and B represent a tradeoff. Specifically, treatment B requires a sharper drop in quality of life immediately in return for a longer life at a steady level of quality. To compare these treatments, we’d have to account for both the earlier quality of life lost with B versus the later quality of life gained.

In both examples above, we see how the QALY framework allows us to compare treatments that differ in terms of both life expectancy and quality of life.

**QALYs and Affective Forecasting**

While QALY calculations are very useful for summary analyses, it’s important to keep in mind that that are only an abstracted estimate – there are some clear reasons to believe that they don’t fully represent life experience.

One very basic issue is that QALY calculations rely on comparisons of health states. Even if a participant is currently experiencing one of the two health states being compared, he or she still has to forecast the other. For instance, a currently healthy person estimating QALYs for a specific health state has to forecast, or estimate, the experience of that state. Now, we can also survey people currently experiencing the health states we want to assess, but we then have the flipside of the same problem in that this population will have to forecast their own experience of absence of the health state.

The bottom line on what we know about this kind of **affective forecasting** (prediction of future or assessment of past states of emotion, comfort, well-being, etc.) is people are pretty bad at it. Specifically, people seem to have poor abilities to forecast our own utility in various different states. We know how we are feeling at the moment, but it is hard to compare that feeling to states we are not currently experiencing (even if we have experienced those states in the past).

As a simple example, note that dialysis patients estimate the QALY level of their current state around .6 per year while currently healthy participants estimate the QALY level they would experience if (hypothetically, for them) they were on dialysis as .36 per year. That is, imagining dialysis results in a much lower QALY estimate than experiencing dialysis does, on average.

Healthy participants, or more precisely those not experiencing the state being assessed, must imagine lowered health status to estimate QALYs. We generally find that participants in these situations don’t do a good or complete job of anticipating what we call **hedonic adaptation**. Once our health states degrade, we often learn to cope, adjusting to health states over time. When we are asked to imagine a challenging health state that we don’t currently have, we are unlikely to anticipate and think through our own coping and adaptation. Instead, we tend to forecast our initial emotional reaction to a decrement in health status, which is likely to be very negative, and report that.

A very specific effect related to and partially causing neglect of hedonic adaptation is the **focusing illusion**. Essentially, if we are asked to think about some new situation, we focus on the aspects of our lives that are directly changed by that situation. Hence, healthy individual forecasting the impact of renal disease and dialysis will focus on the time they would spend receiving dialysis but not on the hours of the day where they would be free of the treatment.

It’s tempting to say we should only estimate QALYs using people actually experiencing the health state being addressed. I’ll call them “patients.” However, it turns out that this is not a full solution. The ultimate requirement is to put health states into relative terms. So, a patient must estimate the relative QALY of their current situation as compared to a situation without the health burden being assessed. But we know that patients aren’t actually great at forecasting this relative difference, either, even if over their life span they have experience without the health burden and then with it.

One thing that specifically seems to happen with patients is something we call response recalibration. That is, they often equate their current, best day to what perfect health would be like. More generally, it is also not that unusual to get what we might think of as protest votes, where people in what most would call a suboptimal health state, such as being wheelchair dependent, estimate their current QALY level as 1.0. The net result is that patient estimates seem to jump around quite a bit, depending on factors such as how a question is asked. The conclusion, then, is that even patients experiencing a health state are subject to some biases in terms of translating that health state into general QALY terms.

Below, are some data from a paper by Riis et al. (2005 *Journal of Experimental Psychology: General*) using a technique called Ecological Momentary Assessment. Participants were asked to carry around an electronic device (a palm pilot) and record their own, current moods when prompted randomly during the day. Responses were scored as -2.0 to + 2.0 to reflect participant moods from negative to positive. Mood is not the same as overall quality of life of course, but these data illustrate some of the issues with affective forecasting.

Riis et al compared a group of healthy participants to a group of renal patients currently undergoing dialysis. In addition to sampling momentary assessment over a time period and averaging these (see row 1 below), the authors also asked participants to forecast various states. The second row below reflects an estimate of mood for dialysis patients; both groups read a standard scenario describing a patient’s life on dialysis and were asked to estimate their own average mood if their experience was the same as the patient’s. Finally, dialysis patients were also asked to estimate what their own moods would have been during the study period if they had never had kidney problems and never needed dialysis (row 3).

|  |  |  |
| --- | --- | --- |
|  | Healthy Controls | Renal Patients |
| Average ecological momentary assessment | 0.83 | 0.70 |
| Imagine hemodialysis and forecast mood (given a realistic scenario, told to imagine you are that person) | -0.38 | 0.63 |
| Imagine never sick and forecast mood | -- | 1.16 |

Using these data, we can compare what renal patients report as their own mood to what healthy controls estimate their mood would be under dialysis. We see that renal patients, on average, report a momentary mood assessment that averages 0.70. Healthy patients estimate that their own mood if they needed dialysis would average -0.38. This difference is strongly statistically significant indicating that healthy controls overestimate the negative impact of dialysis on mood compared to what renal patients report as their current experience.

Note that while renal patient estimates are directionally slightly lower when they are asked to imagine they are the renal patient in the scenario, the 0.63 mean is not statistically significantly lower than their own average over the week of 0.70. The first two rows of the table therefore suggest that we have a relatively stable average mood estimate from renal patients of around 0.6 to 0.7 and healthy patients substantially underestimate that mood.

This might suggest that if we want to understand the impact of dialysis on mood, we have to ask renal patients. But the third row in the table also presents an interesting contrast. Specifically, when renal patients are asked to estimate their average mood in a healthy state, they estimated significantly more positive mood, with a mean here of 1.16, statistically significantly higher than the healthy group reported moods at .83. This suggests that perhaps renal patients themselves do not realize how much they have adapted to dialysis in terms of their own, daily moods. They think they would be much happier than the controls actually are if they had never had kidney disease. That is, they seem to think that dialysis has a stronger effect on their moods than comparison to healthy controls would suggest. So, if we asked renal patients to forecast how much dialysis would degrade a person’s mood, they are also likely to overestimate the difference[[1]](#footnote-1).

We see these sorts of forecasting errors in many different summary judgments of life, not just mood. And the affective forecasting mechanism I’ve described for you have been tied to QALY responses directly.

Overall, we know we can’t treat QALY estimates as objective truth. Of course, almost no self-report data has that status. And, in making decisions, it’s very useful to understand the direction of likely biases. To the degree there is an objective truth about the impact of disease burden, we do know that healthy controls seem likely to estimate the burden as worse than patients experience that burden in real time.

**Additions and Alternatives to QALY**

One way to deal with affective forecasting issues is to augment participant-based QALY measures with expert judgment. Another method I’ve implicitly suggested above is to try to measure perspectives of those with and without a disease burden, and attempt to see where the middle ground seems to be. A third method is to consider using sensitivity analyses to ensure that large policy bets are not made based on small QALY differences. Overall, relative health state estimates are needed for good, comprehensive decision making but they are also very difficult to estimate with any real precision.

Finally, note that measurement of disease burden is an important public health question addressed through multiple research approaches. An international, European-based group named EuroQol has alternative measures of disease burden that use very clear-cut questions such as agreement or disagreement with the statement “I have some problems in walking about.” These objective measures are correlated with participants’ summary assessments of their own health states. Then, researchers can use regression techniques to estimate the impact of items such as mobility restrictions on overall patient-reported health assessments.

The World Health Organization also has an alternative approach to disease burden named the DALY or disability adjusted life hear. This methodology attempts to estimate standardized disease burden weights for specific disabilities, using expert input from providers and others.

Alternatives to QALY measurements also have pros and cons. One should try to assess the impact of disease burden and the value of improving quality of life. It’s also important to recognize that all of these measures are summary estimates rather than objective truth.

**Summary**

* The ICER framework is very useful for assessment treatment value as long as one takes a comprehensive view of costs and benefits of care.
* Considering benefit measurement specifically, the QALY framework is useful for combining disease burden (or impact on quality of life) with life expectancy.
  + Several concerns with the QALY framework exist. For example, we know respondents have difficulties with affective forecasting (e.g., focusing illusion).
  + The issues with QALY don’t generally imply that we should try to ignore estimating the burden of disease, but rather they imply that we should be cautious in our willingness to take action based on one specific method or estimate.

1. Note that in these mood data, patients seem to be over-estimating the impact of their current state on mood. In other research settings, patients seem to under-estimate this impact. The overarching point, however, is that it is not possible to identify a particular assessment as objective truth. [↑](#footnote-ref-1)