Part V: Cost-Effectiveness Analysis

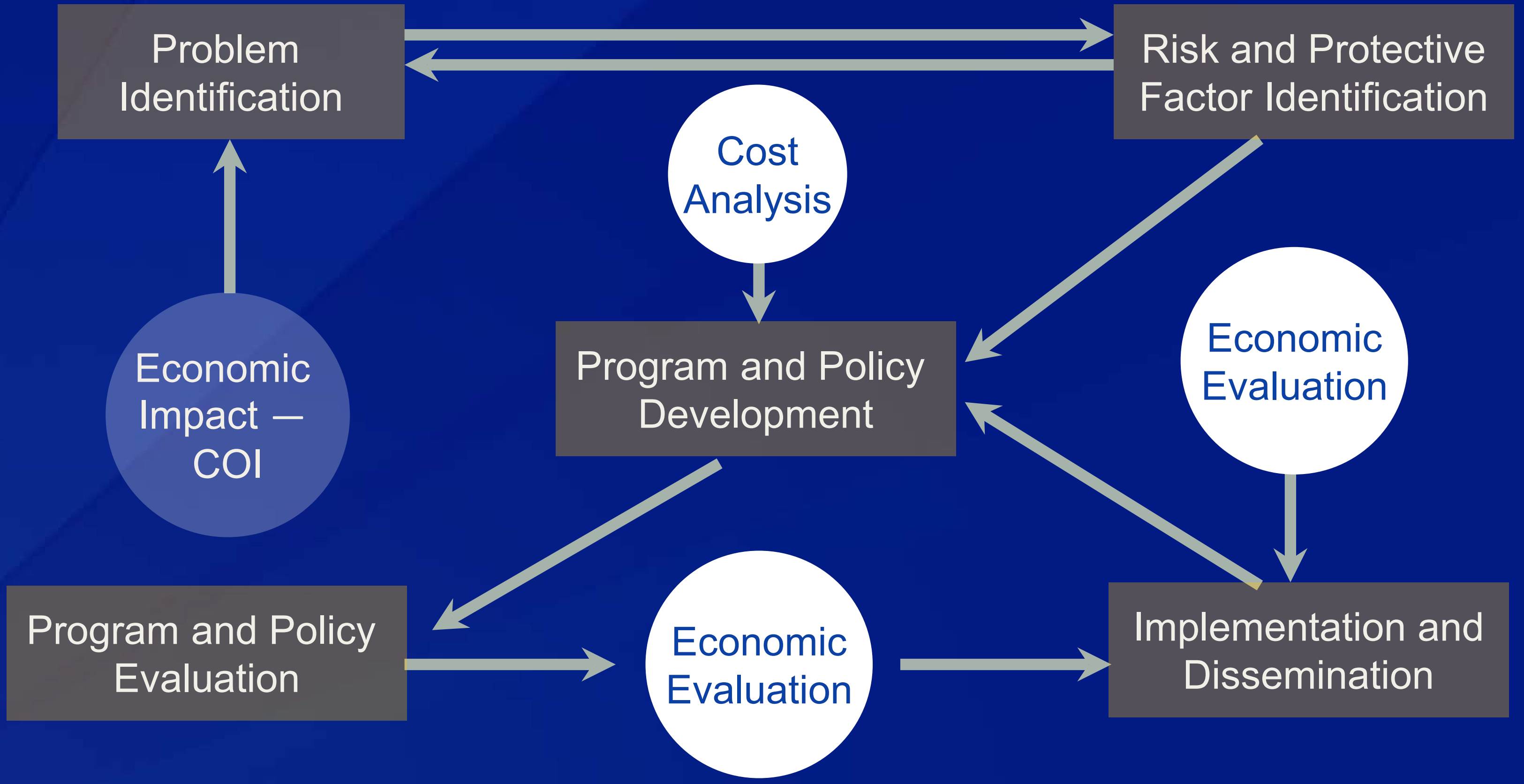
Outcomes in Natural Units:
The Fifth of a Five-Part Series

National Center for Chronic Disease Prevention and Health Promotion
Division for Heart Disease and Stroke Prevention

Slide 1

The last module discusses another type of economic evaluation: cost-effectiveness analysis.

Public Health Model for Prevention



Slide 2

As discussed in the benefit-cost analysis module, economic evaluations are best conducted once a program, policy, or intervention has proven effective but prior to widespread implementation and dissemination. In this way, economic evaluations are typically conducted retrospectively.

However, an economic evaluation is often conducted prospectively, alongside community or clinical trials to ensure efficient allocation of scarce public health resources.

Cost-Effectiveness Analysis (CEA)

- Estimates costs and outcomes of interventions
- Expresses outcomes in natural units (e.g., cases prevented, lives saved)
- Compares results with other interventions affecting the same outcome

Slide 3

As with benefit-cost analysis, a cost-effectiveness analysis compares an intervention’s costs to its outcomes.

Unlike a benefit-cost analysis, a cost-effectiveness analysis expresses outcomes in natural health units, such as the number of cardiovascular disease cases prevented or the number of lives saved, instead of converting outcomes to dollars.

Because of this major difference, cost-effectiveness analysis must be conducted with interventions or programs that impact the same health outcome. For example, you could compare two programs designed to prevent overweight or obesity, where one program focuses on physical activity and the other focuses on nutrition.

CEA — Summary Measures

- Where net costs = program costs A – COI averted

Average Cost-Effectiveness Ratio	Incremental Cost-Effectiveness Ratio
Net Costs_A Net Effects_A	(Net Costs_B – Net Costs_A) (Net Effects_B – Net Effects_A)

Slide 4

The summary measure in cost-effectiveness analysis is the ratio of net programmatic costs divided by net program effects. Programmatic costs are program costs minus the cost of illness averted by the program.

Cost-effectiveness ratios can be an average. One intervention at a time is assessed in terms of net costs divided by net effects.

Two or more programs affecting the same health outcome can be compared in terms of incremental net costs of one program compared to another, divided by incremental net effects of one program compared to another.

Quantify Outcomes — CEA

- Intermediate outcomes
  - Increased physical activity
  - Decreased blood pressure
- Final outcomes
  - Heart disease cases prevented
  - Lives or life years saved

Slide 5

Outcomes or effects included in cost-effectiveness analysis can be defined narrowly or broadly, although broad definitions are preferred for decisions bearing on public policy.

Narrowly defined effects include those that are intermediate in nature and that may be easier to capture, such as immediate increases in physical activity or decreases in blood pressure associated with a hypertension intervention.

Broadly defined effects are those that are more final and further removed, such as cases of heart disease prevented, lives saved, or years of life gained.

These broad outcomes are more appealing in terms of effectiveness goals for a hypertension intervention. However, you may only have intermediate outcomes to work with unless you can follow intervention participants over time or find good epidemiologic evidence linking intermediate to final outcomes.

CEA Caveat

- Outcomes cannot be combined; they must be considered separately. Consider one or two of the most important measures.
- Number of summary measures depends on number of outcomes chosen
  - If A and B are the most important, then
    - Cost/outcome A
    - Cost/outcome B
- Translation for policy-makers can be difficult.

Slide 6

A major caveat in conducting cost-effectiveness analysis is that outcomes in natural units cannot be combined and must be considered separately.

For example, a physical activity program may have two intended effects: lowering blood pressure and decreasing body mass index. Because these two effects can't be combined in a cost-effectiveness analysis, the summary measure for the analysis would be cost per 1 percent reduction in blood pressure and cost per 1 percent decrease in body mass index. However, the cost in these two summary measures is the same, so the ratios are somewhat misleading. This makes cost-effectiveness ratios using natural units difficult for policy-makers to translate.

Cost-Utility Analysis — CUA

- Compares costs and benefits, where benefits = number of life years saved, adjusted for loss of quality
- Combines length and quality of life
- Compares disparate outcomes in terms of utility
  - Quality-adjusted life years (QALYs)
  - Disability-adjusted life years (DALYs)
- Derives a ratio of cost per health outcome
  - $/QALY or $/DALY

Slide 7

One method for dealing with the problem of multiple outcomes, particularly if there are multiple health outcomes, is to conduct a cost-utility analysis. In this type of analysis, outcomes are expressed as a health index. This combines all health outcomes associated with an intervention in terms of increases in length of life and quality of life.

Length of life adjusted by quality of life is known as a quality-adjusted life year, sometimes referred to as a disability-adjusted life year.

In a cost-utility analysis, you could compare interventions that affect different health outcomes by using a quality-adjusted life year—for example, when comparing interventions that affect obesity, nutritional outcomes, and cardiovascular disease.

The summary measure in a cost-utility analysis is cost per quality-adjusted life year or cost per disability-adjusted life year.

When Is CUA Used?

- When quality of life is *the* important outcome
- When the program affects both morbidity and mortality
- When the programs being compared have a wide range of outcomes
- When one of the programs being compared has already been evaluated using CUA

Slide 8

Cost-utility analysis is used when quality of life, rather than length of life, is the most important effect of the intervention.

For example, a cost-utility analysis of a cardiac rehabilitation program might focus on improved quality of life versus the cardiac rehab's influence on the length of life.

Cost-utility analysis is also used when the program affects both morbidity and mortality outcomes. An example is emergency medical services' pre-hospital stroke care, which has long-term effects on recovery and disability.

Cost-utility analysis can be used when comparing interventions that affect different health outcomes, like cancer versus cardiovascular disease prevention.

Finally, cost-utility analysis should be used when comparing results to other studies that also employ cost-utility analysis as the economic evaluation methodology.

Quantify Benefits — CUA

- Utilities, or preference weights, are
  - A quantitative approach for describing preferences for quality of life
  - Typically based on a 0-to-1 scale, where
    - 0 = death
    - 1 = perfect health

Slide 9

Utilities, or preference weights, are a way to quantitatively describe consumer preferences for good quality of life and a subjective measure of the usefulness that results from being in different health states.

Because utilities are quantitative, they are measurable and analyzable. They're typically based on a 0-to-1 scale, where 0 is considered death and 1 is considered perfect health.
To quantify benefits in a cost-utility analysis—that is, to derive a quality-adjusted life year—you need to know the intervention's effect on length of life and quality of life. Data on length of life may be readily accessible from epidemiologic literature.

Effects on quality of life, however, are theoretically derived from individuals directly as their preference weights, or utilities, for the health state under consideration. For example, what is the preference for having a body mass index above 35 versus having one between 25 and 35?

Utilities: Direct Measurement

Response Method	Question Type	
	Certainty (“value”)	Uncertainty (“utility”)
Scale	Rating Scale	
Choice	Time Trade-Off	Standard Gamble

Slide 10

There are a number of ways to directly elicit utilities. There are methods that rely on a specific response method, such as scale versus choice, and methods that rely on a specific type of questioning format, such as asking about certain events versus uncertain events.

Theoretically, to be considered an economic “utility,” the response method must be a choice and the questioning format must include an uncertainty. Therefore, the only correct way to derive utilities for health states is the standard gamble approach, although other approaches are popular in the literature.

Standard Gamble Approach

- The gold standard in utility elicitation
  - Respondent must choose between two options:
    - Living with below-optimal health
    - A lottery between two uncertain health states; for example…
      - Good health with probability x
      - Bad outcome with probability $(1 – x)$
- Limitations
  - Time-consuming
  - Cognitively difficult for many people

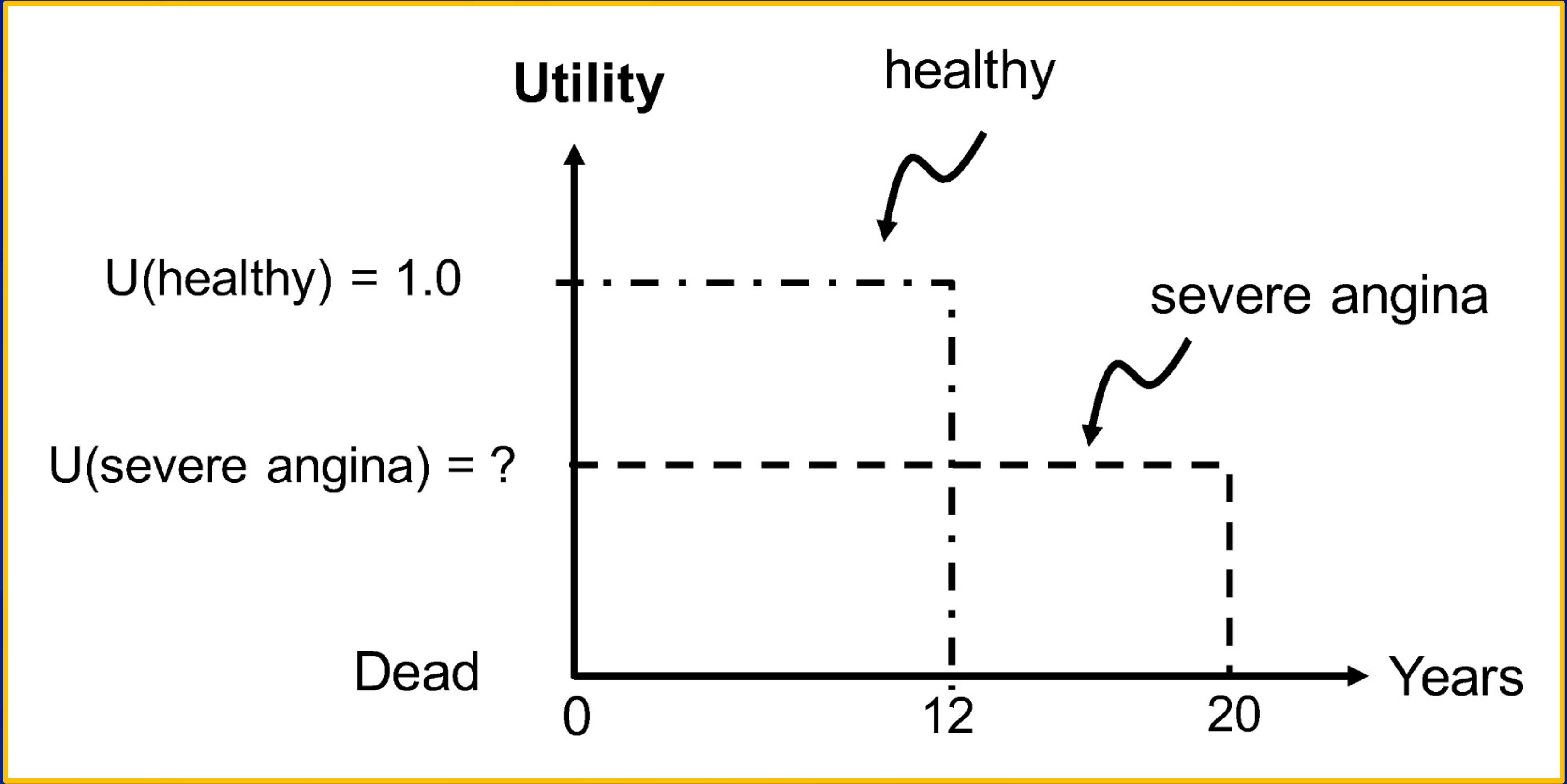
Slide 11

The standard gamble approach is based on the conceptual framework for examining decisions under uncertainty. The respondent is given a choice between a less-than-optimal health state— for example, having a body mass index above 35—and a lottery between two uncertain health states. The two uncertain health states are often perfect health and death and can be valued as 1 and 0, respectively. The two uncertain health states don’t have to include perfect health and death. The only requirement is that the certain health state be in between the two outcomes associated with the gamble.

In this setup, the respondent is asked something like this: Imagine you have a body mass index above 35, with no other adverse health outcomes. Now suppose there’s a surgery available to you that would reduce your body mass index to a perfect level, thus giving you perfect health. However, there’s a probability of death associated with the surgery. How low does the probability of death have to be for you to be indifferent between your certain health, with a body mass index above 35, and the gamble of taking the surgery, which could lead to death or perfect health?

The probability, or *p* value, derived from this scenario reflects the utility for the certain health state under consideration—in this case, body mass index above 35.

Time Trade-Off



Slide 12

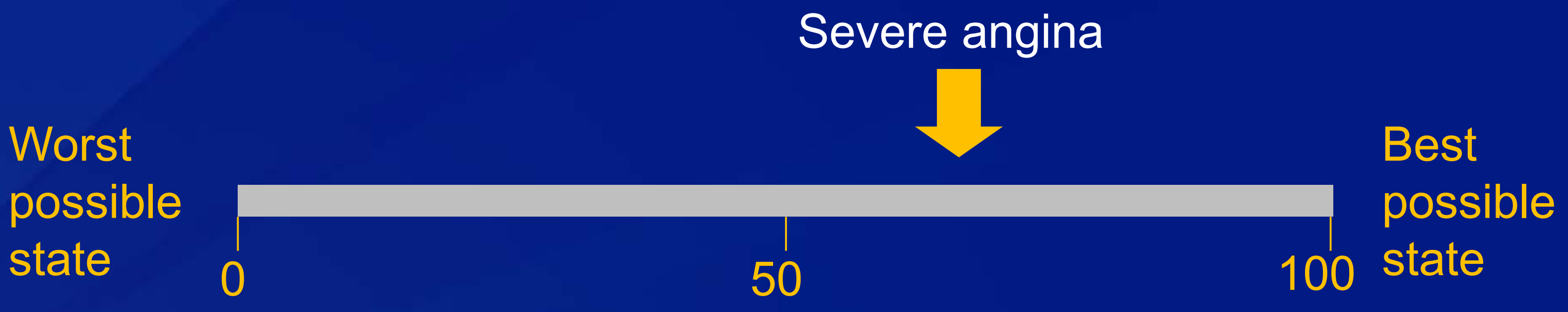
Another way to directly elicit utilities is the time trade-off method, which was developed as an alternative to the standard gamble. This method is used primarily in health research. The respondent is offered a choice between two alternatives of certainty. The goal is to find the point where the person becomes indifferent between the two alternatives.

Here's the setup: Imagine that your remaining life expectancy is 20 years and you have severe angina. How much of your remaining life expectancy would you give up to eliminate your severe angina so that you have perfect health?

The number of years you would give up, divided by the remaining life expectancy and subtracted from 1, represents the utility associated with severe angina.

Rating Scale

- Does not involve trade-offs; not a valid measure of utilities
- Simple, easy to administer



Slide 13

Finally, the rating scale is the most common approach to directly eliciting utilities. This involves ranking alternatives and then placing them on an ordinal scale. For example, alternatives might include perfect health, mild angina, severe angina, and death. This example uses a visual analog scale, which is typically horizontal.

There are a couple of advantages of this approach. The cognitive burden is lower than with other techniques, and people are familiar with the technique. There are several disadvantages, however.

First is the anchoring effect. What is set as the best possible state and the worst possible state is subjective, creating an indexing problem.

In addition, we can’t make any interpretations about the numbers themselves, such as 88 versus 60, because of the ordinal scale.

Furthermore, people have an aversion to the ends of the scale, so they treat the middle of the scale as one scale and the ends of the scale as another scale.

There are also context effects. Ranking and scoring depend not just on the states themselves, but also on the states being compared.

Finally, this approach is based on conditions of certainty and not really tied to utilities or the theoretical foundation on which cost-effectiveness analysis is based.

Where to Get QALY Weights

Source	Examples	Disadvantages
Literature	• Individual studies • CUA databases	• Lack of comparability
Indirect measures	• Beaver Dam study, QWB • Joint U.S.-Canadian health survey included HUI • MEPS included EQ-5D US	• Only common diseases • No severity levels
Direct measures	• Expert panel • Special sample survey	• Expense • Time • Representation

Slide 14

In addition to directly eliciting utilities, there are published preference weights in the literature from individual studies. Compendia of weights are available online at the Tufts Medical Center Web site.

The disadvantage of using weights derived from other studies is comparability. It could be that weights are derived from different populations, for slightly different health outcomes, et cetera.

As an alternative, there are widely available, indirect elicitation tools that involve people classifying their health states based on a number of health domains—such as physical functioning, role, social, and emotional—then applying directly elicited preference weights. Many such tools are available, sometimes for a small fee. The disadvantages of these tools are that their weights may be derived from dissimilar populations, they may not have included the same health outcome you are considering, and they may not have included the same severity levels.

The direct measures we discussed should be elicited from general populations, but expert panels or special disease-specific samples are often used. Major disadvantages are costs, time to collect, and representativeness outside your study.

QALY Weights for Chronic Diseases

- Data from MEPS, 2000–2002
- Regression methods used to estimate disutility for 95 ICD-9 codes, controlling for age, sex, comorbidity, race/ethnicity, income, education
- Results

Marginal disutilities

389	Hypertension	–0.0250
410	Acute MI	–0.0409
427	Dysrhythmia	–0.0190
428	Heart failure	–0.0635

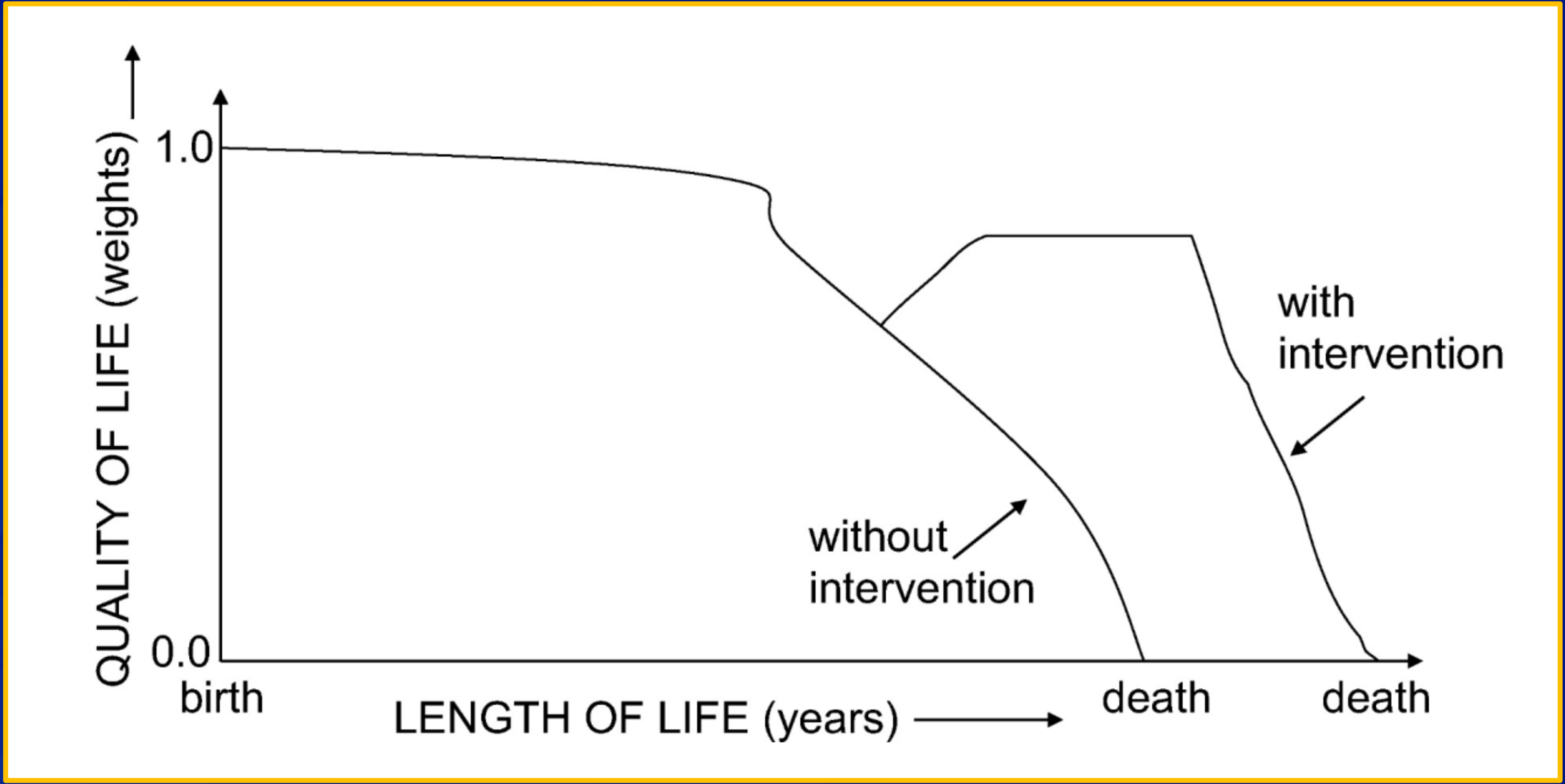
Slide 15

Here's an example of an indirect utility elicitation tool using the EuroQol 5 dimension scale included in the Medical Expenditure Panel Survey for a few years in the early 2000s. Examples of decreases in utilities, or disutilities, are shown for a number of chronic diseases.

Here's how you can interpret these results: Imagine a person with chronic hypertension, with a remaining life expectancy of 20 years. You could say that the person has a quality-adjusted life expectancy of 19 years and 6 months—or a loss of 6 months in quality-adjusted life expectancy. This is derived by multiplying .025 by life expectancy to get .5 years, or 6 months.

Valuation of Benefits in a CEA

Combining Length of Life with Quality of Life



Slide 16

Once the utilities are determined for the effects of the intervention, you can compare the difference in quality of life and length of life between the intervention and no intervention in this example.

CEA Example

- Evaluation of WISEWOMAN program
  - Program cost $440 per 1% reduction in cardiovascular disease risk
  - Cost-effectiveness ratio estimates
    - $4,400 per life year with optimistic assumptions
    - $15,300 per life year with more realistic, long-term outcomes
    - $133,500 per life year if changes are assumed to last only 1 year and only for those with follow-up
- Conclusion: Cost-effectiveness depends on long-term outcomes, which are not known.

Finkelstein EA, Khavjou O, Will JC. Cost-effectiveness of WISEWOMAN, a program aimed at reducing heart disease risk among low-income women. J Womens Health. 2006;15(4):379–89.

Slide 17

Here's an example of a cost-effectiveness analysis of the WISEWOMAN program. The unit of effectiveness was reduction in cardiovascular disease risk, which was then translated, based on epidemiologic evidence in the literature, to life years gained. The uncertainty in the analysis was how long the changes in the cardiovascular disease risk were assumed to last, thus affecting life years saved and costs per life years saved.

The program was assessed in relation to itself, not compared to other interventions, which produced an assessment of the average cost-effectiveness ratio. As a result, the authors found that the program cost 4,400 dollars per life year gained under the most optimal assumptions, which would be changes in cardiovascular disease risk assumed to last a lifetime.

But when more realistic, longer-term outcomes were evaluated, the costs increased to 15,300 dollars per life year saved. These changes in cardiovascular disease risk were assumed to last for the lifetime of only 24 percent of the participants.

Costs were more than 133,000 dollars per life year saved when the cardiovascular disease risk changes were assumed to last only one year and not longer.

This study shows the importance of having longer-term, final outcomes in the cost-effectiveness analysis.

Interpreting Cost-Effectiveness Ratios

- Some analyses assume that only interventions with ratios below $50,000 or $100,000 are “cost-effective.”
  - The U.K. uses a threshold of £30,000 per QALY
- These are arbitrary thresholds.
  - Round numbers
  - Not adjusted for inflation

Slide 18

Unlike in benefit-cost analysis, where the summary measures are objective, cost-effectiveness analysis results in a *subjective* summary measure. The policy-maker must still determine the threshold below which an intervention is considered cost-effective and above which an intervention is not considered cost-effective.

Some arbitrary thresholds have been set in both the United States and the United Kingdom, but there is still some controversy about these thresholds, particularly because the cost-effectiveness ratios haven't been adjusted for inflation.

One way to determine threshold values is to compare cost-effectiveness ratios to ratios published in the literature or to ratios for interventions that are commonly accepted as good practice.

CEA Ratios Example

Cost-saving | Cost-effective (ICER <$50,000) | Intermediate value (ICER ≥$50,000 and <$150,000) | Low value (ICER ≥150,000)

Groups Without Prior Cardiovascular Disease	Men			Women		
Age	Stage 2 hypertension	Stage 1 hypertension, diabetes or CKD	Stage 1 hypertension, no diabetes or CKD	Stage 2 hypertension	Stage 1 hypertension, diabetes or CKD	Stage 1 hypertension, no diabetes or CKD
33–44	Cost-saving	$13,000	$40,000	$26,000	$125,000	$181,000
45–59	Cost-saving	Cost-saving	Cost-saving	Cost-saving	$16,000	$22,000
60–74	Cost-saving	Cost-saving	Cost-saving	Cost-saving	$3,000	$7,000

Figure 1. Projected Average Cost-Effectiveness of Full Implementation of the 2014 Guidelines for Hypertension Treatment in Patients Without Cardiovascular Disease, According to Sex, Age, Hypertension Stage, and Status with Respect to Diabetes and Chronic Kidney Disease

Treatment of each group is compared with the strategy outlined in the previous incremental step in the study model for patients in the same age category. Patients with stage 2 hypertension who do not have cardiovascular disease are compared with patients with cardiovascular disease. CKD denotes chronic kidney disease; ICER, incremental cost-effectiveness ratio.

Moran AE, Odden MC, Thanataveerat A, Tzong KY, Rasmussen PW, Guzman D, et al. Cost-effectiveness of hypertension therapy according to 2014 guidelines. N Engl J Med. 2015;372(5):447–55.

Slide 19

Here is an example from a study analyzing the potential cost effectiveness of treating all hypertension patients according to new 2014 guidelines, with CE ratios for each of the different patient groups. As you can see, the authors have defined what they consider cost-effective for this study and ranked each group according to these defined cut-offs. In this case, less than 50,000 dollars per QALY was considered cost effective, which is a commonly used threshold in the literature.

More on Cost-Effectiveness Ratios

- A continuum, not a decision rule
- Low ratios (< $30,000/QALY) are a good value
- High ratios (> $300,000/QALY) are seldom a good value
- Intermediate ratios are difficult to assess
- Different standards for prevention vs. treatment and clinical vs. population-based interventions

Slide 20

We might think of the cost-effectiveness ratio on a continuum but without an actual rule for policy-making. There are some ranges within which an intervention is clearly a good value, and other ranges within which an intervention is clearly not. It's the intermediate cost-effectiveness ratios that require some subjectivity on the part of policy-makers.

Furthermore, a different set of standards seems to apply to policy-making in the treatment or clinical world compared to the prevention or population-based health world.

But this discrepancy is due in part to the newness of economic evaluations to prevention and public health. Much of the acceptance of economic evaluations for informing policy-making and standardizing practices comes from education modules like this that introduce the concepts to the field. Much has been done in public health since the early 1990s, but there's still a long way to go.

Lessons from Research

- Paper by Grosse, Teutsch, and Haddix covered…
  - Birth defects prevention
  - Newborn screening
  - Cervical cancer screening
  - Childhood immunization
  - Blood product safety
- Findings:
  - Economic evaluation not consistently informing policy
  - Missed opportunities
  - No evidence of consistent threshold for cost effectiveness

Grosse SD, Teutsch SM, Haddix AC. Lessons from cost-effectiveness research for United States public health policy. Annu Rev Publ Health. 2007;28:365–91.

Slide 21

Grosse and colleagues wrote a paper which specifically discussed the use of economic evaluations to inform public policy. The authors found no consistent use of economic evaluations to inform public policy in the United States. The same cannot be said in the United Kingdom, where economic evaluations are part of how the National Health Service determines which benefits are covered.

The authors also found many missed opportunities and no clear thresholds for cost-effectiveness analysis in whatever policies were informed by economic evaluations.

DALYs

- Developed to quantify burden of disease and disability
- Applications:
  - Used primarily for burden of disease and injury estimates
  - Most CUAs in the U.S.; Europe uses QALYs
  - CUAs in resource-poor countries use DALYs, due to World Bank and WHO guidance
- DALY weights:
  - Inverted scale: 0 = health, 1 = death
  - Estimated for diseases or injuries and disabling sequelae
  - Based on preferences of experts, not individuals

Slide 22

Disability-adjusted life years are another outcome measure that can be used in cost-utility analysis.

Disability-adjusted life years were developed in the international community primarily to measure disease and injury burden and to allow comparable estimates of these burden measures across countries.

The disability-adjusted life year weights are slightly different from the quality-adjusted life year weights, with an inverted scale of 0 referring to perfect health, or no disabilities, and 1 referring to death, or 100 percent disabled.

DALY Methods

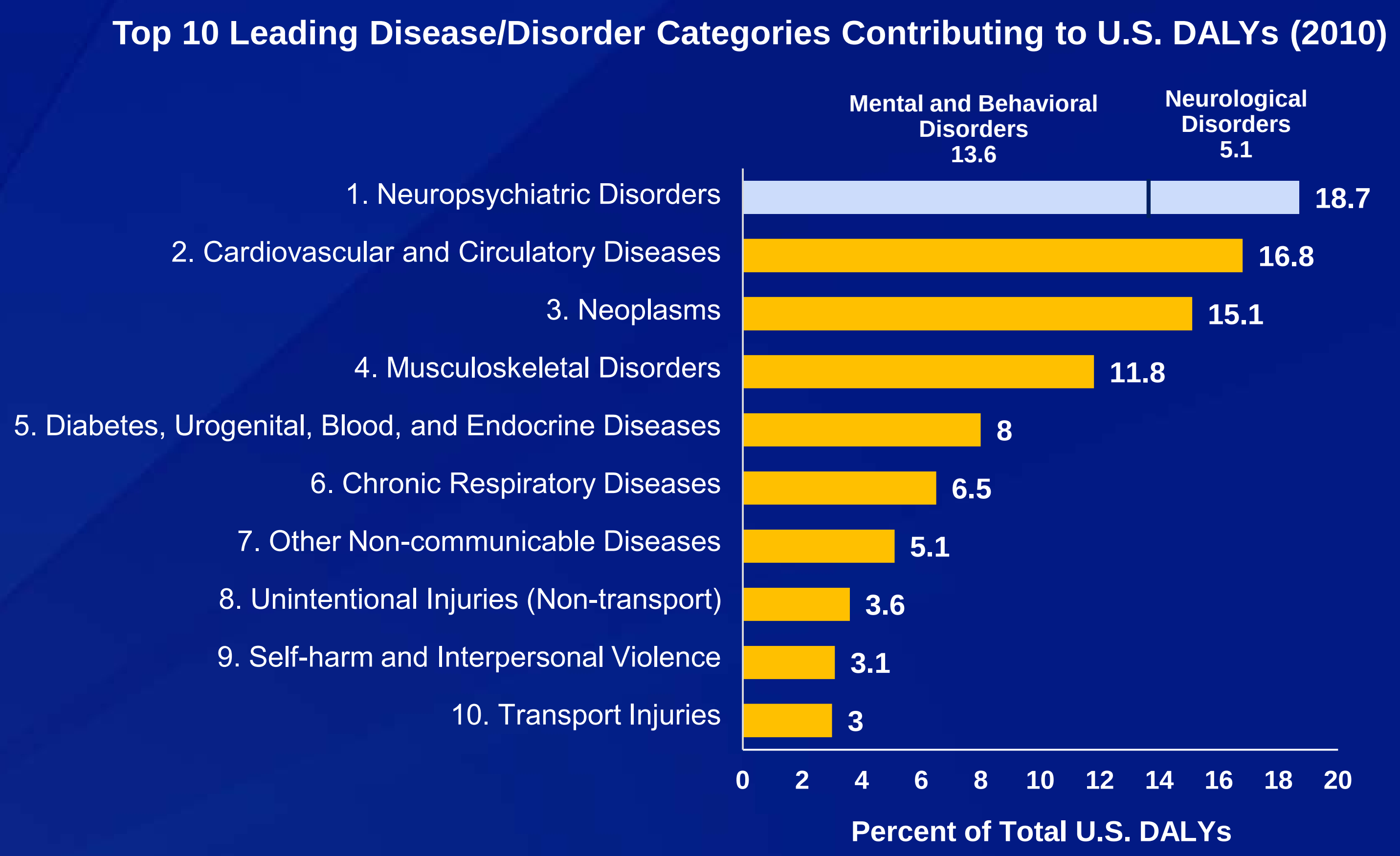
- Combination of…
  - Years of life lost (YLL)
  - Years of life lived with disability (YLD)
- DALY weights
  - Person trade-off (PTO) exercises:
    - Experts asked to trade off numbers of people to keep alive with different conditions
    - Measures relative desirability of conditions, not disability (activity limitations)

Slide 23

Disability-adjusted life years are derived from the estimates of years of life lost—which is a common metric to measure burden of disease internationally—and years of life lived with a disability. It's essentially the same thing as quality-adjusted life years in that life expectancy, in life years, is adjusted for the number of years living with a disability.

Disability-adjusted life year weights are derived differently than quality-adjusted life year weights. Instead of using the standard gamble, time trade-off, or rating scale approaches with a general population sample, experts are asked to trade off numbers of people to keep alive with certain conditions.

U.S. Burden of Disease and Injury



National Institute of Mental Health. Top 10 leading disease/disorder categories contributing to U.S. DALYs website. Available from: http://www.nimh.nih.gov/health/statistics/disability/file_148328.pdf.

Slide 24

Here are the top ten disease categories contributing to DALYs in the U.S. from 2010. As would be expected, CVD is near the top because of the high number of years of life lost, as well as the decreased quality of life it can cause for people living with the disease. But even above CVD is neuropsychiatric disorders; although these disorders don’t contribute much to years of life lost from a disease, the high number of years of life lived with the disorders and the decreased quality of life they cause pushes them to the top contributor of DALYs.

Conclusions

- Economic evaluation is valuable to decision-making and in setting health policy.
- Economic evaluation is both art and science.
- Economic evaluation can help prioritize resources.
- For researchers in public health and prevention, this is an important component of program evaluation.
- Demand for these analyses is growing.

Slide 25

In conclusion, economic evaluation is valuable to decision-making and in setting health policy.

Economic evaluation is both art and science, and it can be used to help prioritize resources for the most effective strategies. It assumes evidence and evidence-based decisions. For researchers in public health and prevention, this is an important component of program evaluation that should be considered because the demand for these evaluations is growing.

Acknowledgments

This presentation was developed by the Division for Heart Disease and Stroke Prevention (DHDSP) at the Centers for Disease Control and Prevention (CDC) under the leadership and guidance of:

Rashon Lane, Applied Research and Evaluation Branch, and

Angela Soyemi, Communications Team

In collaboration with presentation author:

Phaedra S. Corso, PhD, MPA, Department of Health Policy and Management, University of Georgia College of Public Health

CDC's National Center for Health Marketing was integral in the development of this presentation. We thank:

Anatavia M. Benson, **Charlotte U. Duggan,** and **Thomas G. Race**

This presentation was updated in August 2016.

For more information please contact Centers for Disease Control and Prevention

1600 Clifton Road NE, Atlanta, GA 30333
Telephone, 1-800-CDC-INFO (232-4636)/TTY: 1-888-232-6348
E-mail: cdcinfo@cdc.gov Web: www.cdc.gov

National Center for Chronic Disease Prevention and Health Promotion
Division for Heart Disease and Stroke Prevention