SIXTEENTH ANNUAL
HERBERT LOURIE MEMORIAL LECTURE ON HEALTH POLICY

Spending Health Care Dollars Wisely: Can Cost-Effectiveness Analysis Help?

Milton Weinstein

No. 30/2005

Milton Weinstein, the Henry J. Kaiser Professor of Health Policy and Management and Biostatistics at the Harvard School of Public Health, and Professor of Medicine at the Harvard Medical School, is an internationally-known expert on the methods of cost-effectiveness analysis in health care. He is engaged in numerous research projects concerned broadly with issues of resource allocation and decision making in health policy. He directs the Program on Economic Evaluation of Medical Technology, a unit in the Harvard Center for Risk Analysis. He is principal co-investigator for methodology in the Cost-Effectiveness of Preventing AIDS Complications (CEPAC) research group, whose purpose is to evaluate the health and economic consequences of alternative policies for the identification and treatment of persons with HIV infection in the US and the developing world. He co-directed a similar modeling effort in coronary heart disease, which has been used to estimate the cost-effectiveness of national guidelines for serum cholesterol reduction; to evaluate the costs and consequences of alternative drug therapies for hypertension; to predict changes in life expectancy achievable by prevention of coronary heart disease; and to assess the relative contributions of prevention and treatment in the decline of coronary heart disease mortality.

His methodological research concerns the adaptation and use of infectious disease transmission models in cost-effectiveness analyses, and the analysis and representation of uncertainty in economic evaluations. Dr. Weinstein was co-chairman of the Panel on Cost-Effectiveness in Health and Medicine and is a member of the Institute of Medicine. He has served on several of IOM's committees, including the Committee on Priorities for New Vaccine Development. He received his Ph.D. in Public Policy from Harvard University in 1973.

Policy Brief

Spending Health Care Dollars Wisely:
Can Cost-Effectiveness Analysis Help?

Milton Weinstein

Spending Health Care Dollars Wisely: Can Cost-Effectiveness Analysis Help?

One question that is rarely asked, at least in polite company, is whether we're getting the most health improvement possible for our money. In other words, are all the things that we do in medicine really worth it? That is where cost-effectiveness comes in. As a nation, we have been unwilling, at least publicly, to look explicitly at the value, in terms of improved health outcome, that we get for our health care dollars. With advances in medical technology putting unsustainable pressure on health care costs, our historical reluctance to measure value for health care may have to change.

I start this brief by describing cost-effectiveness analysis as a method of determining the value, measured in Quality-Adjusted Life Years, of medical technologies as they are applied to treat, diagnose, or prevent various conditions. Based on this information, I then argue that some highly beneficial, low-cost procedures are significantly underutilized, and that other medical technologies may be overutilized based on the amount of health benefit they yield in relation to their cost. Next, I give examples from current research, my own and that of colleagues, illustrating how cost-effectiveness analysis can be used to guide the use of new diagnostic testing technologies (such as DNA or RNA typing of infectious agents or identification of genomic or proteinomic markers in cancer patients).

Measuring Value for Health Care Dollars

In last year's Lourie Lecture, David Cutler, my colleague from Harvard University, argued that health care is good value *on average.* He asked the question, "Are the benefits of health care worth what we pay for it?" and his answer was generally "Yes, the benefits of health care are worth what we pay for it." But that is not the same thing as saying that

all health care is good value. I submit that some health care is better value than the average and some health care is worse. Even though the average may be acceptable, there are some medical practices that are not worth the money we pay for them. Cost-effectiveness analysis is a method of determining the value of health care treatment.

Dollars

By money we mean the net resources consumed in providing the intervention, measured in dollars. This is relatively easy to determine, although not as easy as you might think. For now let's just pretend that it's relatively easy to measure the resources consumed in providing a health care intervention: the cost of the hospitalization, the physicians, the devices, the drugs, and all the rest of it. That's the money part.

Quality-Adjusted Life Years (QALYs)

Value in health care, I submit, is best measured in terms of improvements in health outcomes, which may be increases in life expectancy or improvements in the quality of life. First, we assign relative weights to different qualities of life, ranging from perfectly healthy, at the top of the scale, or 1, to dead, at the bottom of the scale, or 0. Between those extremes are various levels of health, ranging from maybe having a little cold, which could be a .99, to having a very serious disability, such as after a stroke or possibly the end stages of cancer.

Where do these weights come from? They derive from what people say they want. For example: How much relative value do people place on having relief from chest pain relative to having relief from fever or upper GI symptoms? We survey people in the community and ask them what their preferences are for various health conditions; this is integral in measuring health improvement. Then we take these weights that people assign to being in various states of health, multiply them by their durations, add them up, and we get a number of quality-adjusted life years (QALYs). See, for example, Bell et al. 2001.

Measuring Quality-Adjusted Life Years for an Individual

Figure 1 is a schematic of how one would measure the number of quality-adjusted life years lived by an individual who might be a

candidate for a treatment for asymptomatic coronary artery disease, such as a statin drug. Suppose this patient receives an initial treatment, is asymptomatic for some period of time, and then has a setback, starts experiencing chest pain, angina, for some period of time. This angina is certainly worse than being perfectly healthy. So the height of that bar is lower than the 1.0 for the asymptomatic period of time. After some time with angina, the patient undergoes surgery, recovers pretty well, but still has some minor limitations after the surgery. Eventually the patient suffers a major heart attack, then after a few years of being bedridden, dies. What is the number of quality-adjusted life years? It is essentially the total area underneath these bars. Our measure of health improvement, then, is an integrated measure of length and quality of life.

Figure 1. Quality-Adjusted Life Years: Heart Attack Example
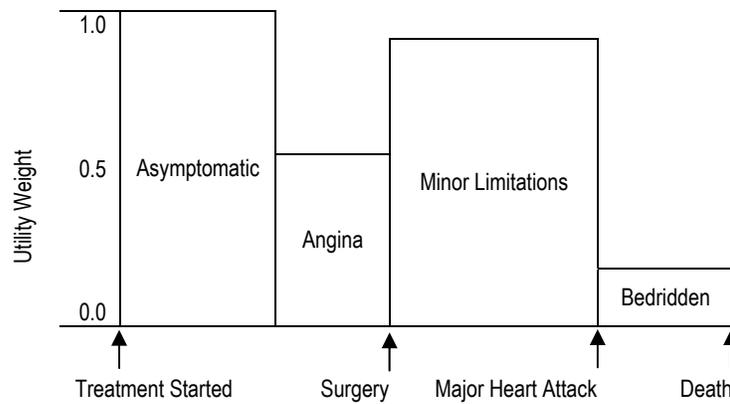
Measuring Value for Money

Now how do we measure the value for money from a medical technology? How can a medical technology or clinical intervention affect costs and health outcomes? First, it can either increase or decrease cost. What do I mean by "decrease cost"? Some technologies may actually save money down the line, which is more than enough to offset the cost that you have to pay for it. The conventional wisdom is that prevention often falls into this cost-saving category, but it turns out that not all prevention is cost-saving. In fact, most prevention really is no more likely to be cost-saving than treatment. It *could* happen that a

treatment is cost-saving, and if it happens that a treatment is cost-saving and also provides an improved health benefit, an improved health benefit in terms of longevity or quality of life or both, then the answer is "Yes, we want this technology." We don't even have to ask the question of value for money because it not only provides health improvement but also saves money. This situation is represented by the upper right-hand box in Figure 2.

Figure 2. Cost-Consequence Space

	Incremental Cost	
	+	**-**
+	?	Yes
-	No	?

Incremental Health Effect

The lower left-hand box in this table is equally clear; it costs money and it harms health outcome. We don't want those. The problem, the battleground where we can't afford to do everything for everyone, is located in the boxes with the question marks, and particularly the one in the upper left-hand corner. This is where most of health care is: it costs money and it provides a benefit. How can we evaluate these kinds of medical technologies?

The Cost-Effectiveness Ratio: QALYs per $1 Million

One measure is literally value for money. How many quality-adjusted life years does this technology buy for each dollar we spend on it?

$$\frac{\text{Net gain in health outcome (QALY)}}{\text{Net increase in health care cost (\$)}} = \text{Value for money}$$

It is the ratio of the quality-adjusted life years gained, the gain in health outcome, relative to the increase in health care costs. This is what I mean by *value for money.*

4

With that methodology it's possible to make comparisons across medical interventions, in terms of how much health improvement you get for every dollar spent on the technology.

Table 1. How Much Health Improvement (Quality-Adjusted Life Years) Can One Million Dollars Buy?	
Beta-blockade post myocardial infarction (MI), high risk	>200
Statin therapy, prior heart disease	100-200
Beta blockade post MI, low risk Antihypertensive treatment, diastolic blood pressure (DBP)>105 Cervical cancer screening every 4 years Screening sigmoidoscopy every 5 years Combination antiretroviral therapy for HIV	50-100
tPA (vs. streptokinase) for heart attack Coronary angiography and revascularization post MI, high risk Antihypertensive treatment, DBP 95-104	20-50
Dialysis for end-stage renal disease Statin therapy, low density lipoprotein (LDL)>190, high risk	**10-20**
Coronary angiography post MI, low risk Statin therapy, LDL 160-190 or >190 and low risk Cervical cancer screening every year (vs. 2 years)	<10
Sources:Goldman et al. (1991), Prosser et al. (2000), Edelson et al. (1990), Frazier et al. (2000), Freedberg et al. (2001), Winkelmayer et al. (2002), Kuntz et al. (1996), Kim et al. (2002), Chapman et al. (2000).	

Table 1 lists various medical procedures. The numbers on the right side tell you how many quality-adjusted life years, on average, this medical technology will buy for every $1 million spent on it. At the top of this particular list is beta blockade—beta blocker drugs used after myocardial infarction—for patients who are at high risk for subsequent cardiac death. This is a very cost-effective intervention; more than 200 quality-adjusted life years are gained for every $1 million spent.

The "Gold Standard" of Twenty QALYs per $1 Million: Treatment of End-Stage Renal Disease

The procedure highlighted near the bottom—dialysis for end-stage renal disease—which is in the category of 10 to 20 quality-adjusted life years gained per $1 million, is an interesting one. In 1972 Congress passed the End-Stage Renal Disease (ESRD) Amendment to Public Law 92-603, which expanded Medicare coverage of persons under age

65 who are disabled to include those with chronic renal failure, thus creating the only virtually universal health coverage program in the United States based solely on a disease. The best treatment for chronic renal failure is a kidney transplant; however, if a transplant isn't available, these Medicare beneficiaries are entitled to dialysis. One of the earliest published cost-effectiveness analyses (Klarman et al. 1968) estimated that dialysis cost roughly $50,000 per year of life gained, or 20 QALYs per $1 million. Surprisingly, that ratio, $50,000 per year of life gained, has had tremendous durability for patients on dialysis. Even as the cost of dialysis went up, the benefits also went up as immunosuppressant drugs were introduced and proved effective (Winkelmayer et al. 2002). This benchmark of 20 QALYs per $1 million is frequently used as a benchmark for society's willingness to pay for another life-year—although lower thresholds such as 10 QALYs per $1 million have also been proposed (Cutler and Richardson 1998). Presumably we ought to be willing to pay for everything above it on the list as well. Indeed, there are several interventions above it on the list:

- *Beta blockade* post-MI is more cost-effective than dialysis, not only for high-risk patients but even for some low-risk patients.

- On the other hand, *coronary angiography after a heart attack,* with the purpose of identifying patients who can benefit from coronary revascularization, either surgery or angioplasty or stent implants, can buy QALYs at about 20-50 per $1 million for high-risk patients, but for low-risk patients it falls near the bottom of the list at less than 10 QALYs per $1 million.

- *Cervical cancer screening* every four years is a very cost-effective thing to do. It is known to be beneficial and it is good value for money, 50-100 QALYs gained per $1 million. But doing a pap smear every year instead of every 2 years gains fewer than 10, maybe a lot fewer than 10, QALYs per $1 million.

These examples show that it is not really correct to say that a technology either is or is not cost-effective. Rather, it depends on how you are going to use it, which patients are going to get it, how often you are going to do it. *It is the way you use technologies that determines how cost-effective they are.*

6

Cost-Effectiveness Ratios: Dollars per QALY

It is much more traditional to express these ratios upside-down, to measure dollars per QALY instead of measuring quality-adjusted life years per $1 million. The medical literature contains numerous studies that report cost-effectiveness ratios in terms of dollars per quality-adjusted life year gained. These tend to run in the thousands up to the tens of thousands, even up into the hundreds of thousands of dollars per QALY.

Reallocating Current Resources to Highly Cost-Effective, Widely Applicable Technologies

There is a wide range, maybe a factor of 100, between the *most* and the *least* cost-effective medical interventions that are widely used (Table 1). And there is surprisingly little difference in the penetration of these technologies at the top of the list versus the penetration of these technologies at the bottom of the list into patient care. So there is a lot of room for reallocation of resources. We could improve our health outcomes as a community by taking some of the money that's going into the technologies at the bottom and reallocating them to the technologies at the top, which are being underutilized.

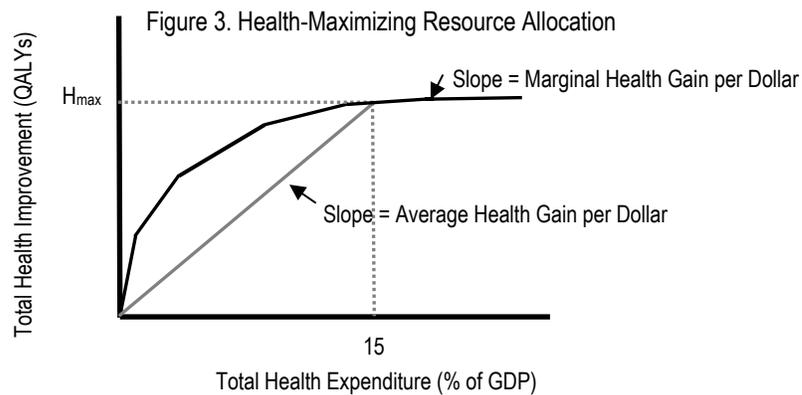Figure 3. Health-Maximizing Resource Allocation

Figure 3 illustrates the total *possible* health improvement due to medical care, measured in quality-adjusted life years, for the amount we choose to spend. There is a little hash mark at 15% of the gross domestic product (GDP) to indicate our current level of health care

spending nationwide. The average health care gain per dollar is the slope of the diagonal line: this is the ratio of how much we get to how much we spend, and it is quite steep, representing good value of money on average. But look at the last percentage point or so of our health care dollars; they're really not buying very much. We're getting up to what Alain Enthoven (1980) calls "flat-of-the-curve medicine," where there is little or no health improvement per dollar.
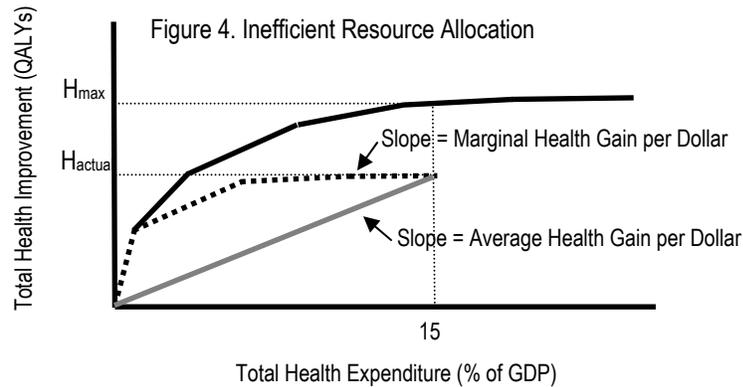
Suppose that as a community we decide that we want our health care resources to be spent in the way that gives us the most possible health improvement as a community. What should we do?

We should make sure that the interventions that give us the most value for money, the most health improvement per expenditure (QALYs per $) are done *first,* in the first part of the graph, where the slope is very steep and there is a lot of health improvement for a little bit of money. This includes beta blockers after MI and childhood immunizations; this is the really good-value-for-money medicine. We would want to make sure that everybody who needs those interventions gets them. So far, we've spent maybe 1/10% of GDP, and we've already gotten a fair amount of health improvement. Then we slide upward along the curve as we move down the list in Table 1 to the interventions that have slightly less value for money. We keep adding technologies until we hit our budget, 15% of the GDP.

At this point, there is still more medical technology available that we're not doing. If we're only willing to spend 15% of the GDP, we're going to have to forego some potential health improvement.

But that isn't how we allocate health care resources in this country. Instead, we tolerate situations where some highly cost-effective medical interventions don't get done at all (Figure 4). Yes, we do the cost-effective interventions that we can't *not* do, such as emergency appendectomies, but we skip over a few cost-effective interventions, maybe because they're not very popular: colonoscopies, for example. People aren't beating down their doctor's doors to get colonoscopies. So we tolerate a situation where only a fraction of the population who could benefit from a colonoscopy to screen for colon cancer actually gets one. We skip over a few things until we reach less cost-effective technologies. Then, when we hit our 15% spending level, we have not

8

maximized the potential health improvement. We're only getting part of the health improvement that we could get.

Figure 4. Inefficient Resource Allocation

Total Health Improvement (QALYs)

H_{max}

Slope = Marginal Health Gain per Dollar

H_{actual}

Slope = Average Health Gain per Dollar

15

Total Health Expenditure (% of GDP)

If, starting at this point, we were to decide to add in all these underutilized but cost-effective medical interventions, we would end up spending even more. So instead we tolerate the fact that we're not getting as much health improvement as we could for a country that spends 15% of the largest gross domestic product in the world on health care. We tolerate less than full utilization of good medical technologies to avoid spending even more than we are on health care.

Underutilized Recommended Medical Technologies

Table 2. Poor Adherence to Quality Indicators: De Facto Cost Containment?	
Type of Care	Percent of Recommended Care Received
preventive	54.9
acute	53.5
chronic	56.1
Mode of Care	
medication	68.6
laboratory testing or imaging	61.7
surgery	56.9
counseling or education	18.3
Overall	54.9
Source: McGlynn et al. 2003, tables 3 and 4.	

9

McGlynn and colleagues (2003) conducted a study on the percentage of recommended care that was received in different categories of care: on average only about 50 to 60 percent of recommended care is actually being delivered. What would it mean to increase that to 100 percent? It would mean more cost. But many of these underutilized services are cost-effective. They're better value for money than a lot of what we're doing. Examples of underutilized services include:

- **Flu vaccine for people over the age of 65.** Studies have shown that this might even save more money than it costs, or at worst might cost $12,000 per QALY gained, which is a very good buy (Coffield 2001, Tables 1,2). Yet only 66 percent of Americans over 65 got flu vaccine in 2001 (CDC 2001, Table 7.1), and despite public education efforts only 74.5 percent of those on Medicare in 2003 received the vaccine (National Committee for Quality Assurance 2004, 36). This proportion was almost certainly lower in 2004 because of the nationwide shortage of flu vaccine.

- **Getting lipids managed after coronary heart disease events such as heart attacks.** Statin drugs are very cost-effective for people who have already had a coronary attack and are at high risk of subsequent events. The cost per quality-adjusted life year is less than $10,000 for most risk groups, and even for lower risk women it is only about $40,000 per QALY (Prosser et al. 2000). And yet in 2003 only about 65 percent of people with commercial insurance or Medicare who could benefit from these medications were getting them, and only 39 percent of those on Medicaid (National Committee for Quality Assurance 2004, Appendix 1).

- **Colorectal cancer screening of people over the age of 50.** Annual fecal occult blood testing (FOBT) is very cost-effective, less than $20,000 per QALY (Frazier et al. 2000). Even colonoscopies or sigmoidoscopies every 5 years cost less than $50,000 per QALY, which is the magic number for kidney dialysis. And yet as of 2003 only 47.4 percent of commercially insured Americans and 49.5 percent of Medicare recipients received FOBT in the measurement year, sigmoidoscopy or double contrast barium enema within the last four years, or a colonoscopy within the last nine years (National Committee for Quality Assurance 2004, 32).

We tolerate this underutilization, and we restrict access to care, so that not only doesn't everybody get the beneficial services, but we make it difficult for some people to get *any* services. Again, the net effect is that we end up spending a large amount of money but get less benefit than we could, and we find that there's a lot that we *are* doing that is less beneficial per dollar than things that we *are not* doing.

Evaluating New Medical Technology

To evaluate a medical technology, we have to decide how it is going to be used, and then compare different uses. Most technologies that were worth developing in the first place are going to be effective, and even cost-effective, for *some* people *some* of the time. But there may be other people or other situations where they are much less cost-effective.

Levels of Implementation

We have to look at the cost-effectiveness, the ratio of the incremental cost to the incremental gain in QALY, for different levels of implementation or adoption of that technology, where the levels of implementation are defined by:

- who gets it (the target population)

- how we deliver it (modality), what particular form of the technology are we going to use, and

- how often we are going to do it (frequency).

Example: Cervical Cancer Screening

These technology, modality, and frequency questions are nicely illustrated by a study that Sue Goldie and colleagues recently did on cervical cancer screening in clinical practice (Kim, Wright, and Goldie 2002).

The major cause of cervical cancer is human papillomaviruses (HPVs), a group of more than 100 known types of viruses (National Cancer Institute 2004). Some HPVs cause warts, or papillomas, which are benign, but others are associated with certain types of cancer. HPVs that are more likely to lead to the development of cancer are labeled

"high-risk," as opposed to "low-risk" viruses that rarely lead to cancer. Even the majority of high-risk HPV infections go away on their own and do not cause cancer.

Both high-risk and low-risk HPVs cause the growth of abnormal cells in the cervix, which can be detected by means of a pap smear. A pap smear removes a small amount of squamous, or epithelial, cells from the surface of the cervix. Among the pap smears performed annually in this country, about 2 million result in findings that are neither unequivocally normal nor cancerous, but are described as "atypical squamous cells of uncertain significance (ASC-US)." It's been a matter of controversy what to do when the findings are equivocal, as in ASC-US. The least costly strategy, which would probably be considered malpractice in some parts of this country, is to ignore those atypical findings, basically treat them as if they were normal pap smears. The standard of care, however, has been to repeat the pap smear until the results are unequivocal.

Another approach to these atypical findings is a colposcopy, which may include freezing the abnormal cells or a cone biopsy, to determine with virtual certainty whether there is a lesion there or not.

In 1999 the FDA approved a new technology, the Digene HPV DNA test, with which it is now possible to detect HPV virus in patients, and to distinguish whether it is a high-risk or a low-risk type. This DNA test is more expensive than a pap smear. Is it worth it?

Kim, Wright, and Goldie (2002) performed a cost-effectiveness analysis of four management strategies for following up ASC-US findings: immediate colposcopy; HPV DNA testing, with a follow-up colposcopy if high-risk HPV types are detected; repeated pap smears; and reclassifying ASC-US as normal, that is, ignoring the equivocal result. They concluded that "reflex HPV DNA testing provides the same or greater life expectancy benefits and is more cost-effective than other management strategies for women diagnosed as having ASC-US." Table 3 summarizes their findings.

These are incremental costs per life year gained, not quality-adjusted, but in cervical cancer the goal is basically to prevent fatal cancer. So cost per life-year pretty much captures what we're interested in (except

12

that all these ratios would be slightly higher if they were expressed as costs per QALY, because the life years gained are in less than perfect health). Compared to doing nothing, doing a pap smear every 5 years and ignoring the atypical findings is a very cost-effective thing to do ($7,100 per life year), but we can do better at only a little bit more cost. Instead of ignoring the atypical findings we could follow them up with an HPV test which, compared to doing nothing, would gain additional life-years at only about $12,000 apiece. The new pap smear technology, using liquid-based methods, is more expensive than the conventional pap smear, but it is also more effective, more sensitive and specific, and it turns out that it is good value to use the liquid-based technology.

Table 3. Incremental Cost-Effectiveness of ASC-US Management Strategies and Screening Intervals	
Strategy	Incremental Cost/LY
Liquid-based, ignore ASC-US, 5 years	$7,100
Conventional, HPV, 5 years	12,100
Liquid-based, HPV, 5 years	20,300
Liquid-based, HPV, 3 years	59,600
Liquid-based, HPV, 2 years	174,200
Liquid-based, HPV, 1 year	794,000
Liquid-based, colposcopy, 1 year	1,800,000
Source: Kim, Wright, and Goldie 2002	

Intervals. We also find that these expensive technologies—the liquid-based pap smear, and the human papilloma virus test—to work up atypical pap smears are very cost-effective as long as we test every five years. If we start thinking about doing it every 3 years, as is now widely recommended, the cost per life year gain is about $60,000, roughly equivalent to kidney dialysis. If we screen more often, say every 2 years or 1 year, the additional gain per dollar spent is very small. Or, as shown by these big numbers, the additional cost per life year gained is very high. What we also conclude is that these new technologies—liquid-based pap smears and the HPV test—are quite cost-effective, but that the annual pap smear is not. High tech wins out over low tech.

Policy has changed, at least in part, because of this study. The American Cancer Society guideline committee that was set up to re-examine cervical cancer screening in the United States in light of these new technologies concluded that the HPV test should be done to work

up atypical pap smear results, but that screening every 3 years, at least until there's an abnormal or atypical finding, is sufficient (Saslow et al. 2002).

Example: Genotypic Antiretroviral Resistance Testing (GART) to Guide HIV Therapy

Human Immunodeficiency Virus (HIV) is the virus that causes AIDS (Acquired Immune Deficiency Syndrome). Left untreated, HIV infection can damage a person's immune system and progress to AIDS (AIDSinfo 2004). HIV is a retrovirus, a type of virus whose genetic information is contained in RNA rather than DNA. The U.S. Department of Health and Human Services (HHS) provides treatment guidelines that recommend a combination of three or more antiretroviral medications in a regimen called Highly Active Antiretroviral Therapy, or HAART. Using HAART has dramatically reduced HIV-related mortality and morbidity: a successful drug regimen reduces the amount of virus so that it becomes undetectable in a patient's blood. However, the initial HAART can fail for several reasons: (1) these drugs may cause negative side effects, some of which may lead patients to discontinue treatment; (2) the HIV can mutate while a person is taking anti-HIV medication; or (3) drug-resistant strains may take over if the virus is not adequately suppressed by medication. Second and subsequent drug regimens are less effective in reducing a patient's viral load, largely because of drug resistance.

My colleagues and I recently did a cost-effectiveness study of a relatively new technology using genotyping for patients with HIV who have failed at least one antiretroviral treatment regimen (Weinstein et al. 2001). A Genotypic Antiretroviral Resistance Test, or GART, determines if the HIV in a particular patient has become resistant to the drugs the patient is currently taking or to the drugs that the patient's physician is considering prescribing next. The test analyzes a sample of the virus from the patient's blood to identify mutations that are associated with resistance to specific drugs. Thus it can help identify the drugs to which these patients are likely to respond, or to which their strain of HIV has not yet developed resistance, and increase the effectiveness of the next anti-HIV regimen. It is an effective technology. But it is also expensive, about $400 in 2001.

14

And GART is only modestly effective. It can increase the suppression rate of the virus in the second- and third-line regimens from 14% to 20%, up to maybe 30%, but that is still not nearly as effective as the first-line regimens. So there is a lot of resistance to adopting this technology.

In our study, we used a model, a so-called state transition decision analytic model: think of it as a "black box" of people with HIV infection. This model synthesizes evidence from clinical trials, from cohort studies such as the Multi-Center AIDS Cohort Study, efficacy of the treatments from clinical trials, and the efficacy of genotype testing from clinical trials. We put all these numbers into our model and out came our cost-effectiveness result, which is that, comparing a world without genotype testing to a world with genotype testing, patients with AIDS can gain about 2.5 months of QALY, and that the cost per quality-adjusted year of life gained is about $16,000. This is less than kidney dialysis, even less than a lot of those very high-value technologies in Table 1. It falls in the same range as treating high blood pressure for people with fairly high diastolic blood pressures, cervical cancer screening even every 4 years, or screening sigmoidoscopy for colon cancer every 5 years.

Basically this study says that if it's worth treating HIV with antiretroviral drugs, it's worth doing the genotype test. To some, that was a surprising conclusion. It shows that expensive technology can be good value, but you have to look at how it's being used, and whom it's being used on.

Example: HER-2 Testing and Trastuzumab Therapy for Metastatic Breast Cancer

Breast cancer is a general term used to describe several different cancers that occur in the breast. One particularly aggressive breast cancer involves a genetic alteration that generates extra copies of the HER-2/neu gene, which sends protein signals to cells to grow and multiply normally. If the cell has extra HER-2/neu genes that produce too much signal protein, the growth signals are out of control, and breast cells reproduce abnormally and form a cancer (see http://www.vysis.com/Herceptin_35577.asp). Approximately 25% of

breast tumors are of this type. The HER-2 gene was linked with breast cancer in 1987.

The drug trastuzumab, produced by Genentech Incorporated under the name Herceptin, is designed to block excess HER-2 protein. It is a monoclonal antibody that was approved by the FDA in 1998 to treat metastatic breast cancer in women whose tumors exhibit HER-2 protein overexpression (extra amounts of protein) or gene amplification (extra copies of the gene). Like many new drugs, it targets the tumor cells that overexpress the HER-2 protein. It is administered in combination with conventional chemotherapy, and randomized controlled trials show that trastuzumab improves the response rate of the tumors and the duration of the response, increases the time to progression of the cancer, and even increases overall survival. It is an effective drug, but it is expensive. Moreover, only 20% to 30% of women with breast cancer can actually benefit from it.

There are two kinds of tests to identify the women who can benefit from Herceptin.

• The immunohistochemistry (IHC) assay, which costs only $85 or so, measures HER-2 protein present on the surface of the tumor cells. The test results are expressed as 0, 1+, 2+ (weakly positive), and 3+ (strongly positive). A test result of 3+ indicates that the patient is HER-2 positive, but results of 2+ or even 1+ may occur in women who are HER-2 positive.

• The fluorescence *in situ* hybridization (FISH) test measures the HER-2/neu abnormality at the DNA level, that is, the number of HER-2/neu genes in the cell nucleus, and it is significantly more accurate than the IHC test. But it costs $400.

Using the less expensive IHC test alone, about 8% of women who could benefit from Herceptin and would be identified as candidates to benefit from Herceptin if they got the FISH test, are missed by the inexpensive test, because there are 8% false negatives. On the other hand, of the women who truly can not benefit, 1.7% would be identified as strongly positive and could get Herceptin unnecessarily and without benefit, based on the IHC test results alone.

Is it worth doing the $400 FISH test instead of the $85 IHC test to avoid missing the 8% of women who could benefit, and also to avoid unnecessarily giving Herceptin to the 1.7% of women who would be identified as positive or strongly positive (even more if you include the weakly positive results), but who really aren't going to benefit?

Our cost-effectiveness study (Elkin et al. 2004) analyzed five test-treat strategies based on results from clinical trials. The comparison strategy was no test and chemotherapy alone. This was compared to four strategies that involved various ways to identify and treat patients with Herceptin: (1) IHC test, Herceptin if the IHC result was 2+; (2) IHC test, Herceptin if 3+, confirm 2+ with FISH and give Herceptin if FISH results were positive; (3) FISH, Herceptin if FISH results were positive; and (4) no test, Herceptin to all patients.

What we found was that doing *anything* is expensive per unit of benefit. This is because of the poor prognosis with metastatic breast cancer, and because the patients most likely to benefit from Herceptin have the poorest prognosis to begin with. Table 4 shows the costs per QALY gained.

Table 4. Costs per QALY Gained for Test-Treat Strategies			
Strategy	Dollars	QALYs	$/QALY
No test, chemotherapy for all	$43,300	1.28	-
IHC test, FISH for 2+, Herceptin for FISH +	$53,700	1.36	$125,000
FISH, Herceptin for +	$54,700	1.37	$145,000
IHC Test, Herceptin for 2+	$57,500	1.36	Dominated
No test, Herceptin for all	$79,200	1.37	Dominated
Notes: (a) The $/QALY is the ratio of the added cost to the added QALYs, comparing each strategy to the next less costly one. For example, the $/QALY for "IHC test, FISH for 2+, Herceptin for FISH + is calculated as ($53,700-$43,300)/(1.36-1.28). Dollar amounts are rounded. (b) *Dominated* means that the strategy is both less effective (fewer QALYs) and more costly than at least one of the other strategies. The last two strategies are dominated by the strategy "FISH, Herceptin for +"; their costs are higher but their QALYs are the same or lower.			
Source: Elkin et al. 2004.			

In contrast with the $15-20 thousand per QALY for AIDS drugs and resistance testing for AIDS treatment, and in contrast with $50,000 per QALY for kidney dialysis, these cost-effectiveness ratios are over

$100,000. In this context, the difference between using the $85 IHC test and the $400 FISH test is very small. You can conclude from this that if you are willing to spend the money to use anything at all, even to use Herceptin in the first place, you should probably be willing to use the expensive test to be sure that you get the drug to the right patients.

Is it worth spending $125,000-$145,000 to gain one QALY for women with metastatic breast cancer? Well, it's hard to say no to that question, but it does raise an interesting, and provocative, issue. A lot of women who should be getting screening mammograms aren't getting them. Maybe if we made sure that all the women who could prevent this cancer by getting mammograms got them, we could do more good than by using these drugs. Of course that raises very difficult ethical issues. Do we want to say "no" to patients with cancer?

The conclusions from this analysis are that if targeted treatment is worth doing, that is, if giving Herceptin is worth doing at all, then so is the more accurate and expensive test. But also that this whole set of interventions, both the testing and the treatment, have a relatively high cost-effectiveness ratio compared to other therapies and screening procedures for breast cancer.

Who Cares About Cost-Effectiveness Analysis?

Government Agencies

You would think that in the United States the government would care. You would think that our representatives in Washington and in our state capitals would care that we're getting the most health value for our money. To a certain extent they do care, but publicly our governments have been reluctant to embrace cost-effectiveness analysis (Neumann 2004).

The Centers for Disease Control has taken a leadership position in using cost-effectiveness analysis in formulating guidelines for vaccination and screening programs, but those are basically recommendations.

You would think that Medicare and the Medicaid agencies in the states would be using cost-effectiveness analysis, but they have pretty much avoided it like the plague. For political reasons Medicare has never

18

used cost-effectiveness analysis, at least openly or officially, and Medicaid agencies have only recently begun to look at it. In fact, as soon as our paper about the cost-effectiveness of genotype resistance testing in HIV was published in the *Annals of Internal Medicine*, I got a call from somebody in the Florida Medicaid office: "We saw your paper. We're trying to get through the state legislature a law that will allow reimbursement for genotype resistance testing. Can you advise us on how to evaluate it?" Because the legislature was interested and was willing to provide reimbursement on a trial basis, the state Medicaid agency was interested in evidence that it would be a good use of state funds.

Virtually the rest of the developed, industrialized world is waist deep in cost-effectiveness analysis. In the United Kingdom, which admittedly has a very different health care system, they have set up an organization called the National Institute for Clinical Excellence, with the beautiful acronym NICE, which requires cost-effectiveness analyses of pharmaceuticals and other technologies prior to a recommendation from NICE to the National Health Service regarding reimbursement. This agency has really been a full employment act for my colleagues in the UK who are doing some wonderful work, both on methodology and application of cost-effectiveness analysis. The National Health Service has actually put some restrictions on reimbursement for new drugs and technologies on the strength of cost-effectiveness analysis.

This is also true of other countries in Europe, in Australia, and in Canada. The Australian government and the Ontario provincial government in Canada were the first to require that reimbursement under their drug benefit programs be guided, and are still guided in part, by cost-effectiveness analysis. It appears that the rest of the world is already paying attention to value for money.

The United States, on the other hand, is doing end runs around the issue by using other mechanisms: restricting access, demanding higher standards of evidence for new technologies, tolerating underutilization, and so on.

Here in this country we rely a lot on the private sector: private health care payers, managed care organizations, and insurers. Bernie Bloom recently surveyed the people who identified themselves as the users of

economic evaluation information in managed care organizations (Bloom 2004). Forty-two percent of them said that they used cost-effectiveness analysis or cost-benefit analysis. There may be others who wouldn't admit it, but some managed care organizations are beginning to dip their toes in the water. They simply have to.

Distorted Incentives

Hospitals

Some hospitals are starting to use, or at least pay attention to, cost-effectiveness analysis, but there are barriers. For a hospital, for example, costs can be offset by revenues, and sometimes the incentives get distorted a bit.

Take, for example, the case of drug-eluting stents. Drug-eluting stents have been shown to reduce the incidence of repeat coronary revascularization procedures, compared to bare metal stents. These drugs, which flow through the arteries, prevent a restenosis of those arteries that would otherwise require either more angioplasty procedures, stents, or bypass operations. Drug-eluting stents are a beneficial procedure that prevent subsequent bypass operations and angioplasties, so they save money. But not for the hospital, because the hospital can get paid to do those coronary bypass operations and angioplasties, and so can the physicians and surgeons. As a result, from the hospital's point of view it's a losing proposition to prevent bypass operations and angioplasties. The incentives are distorted. Even though the costs are reduced, the net revenue, the net income to the hospital, is actually harmed. This is not the fault of the hospitals, but it's the way we've set things up in this country.

Insurers

Insurers tend to have a short-time horizon. If they're for-profit they have to satisfy their shareholders that they've got good earnings this year. Besides, since most of their members are going to leave them pretty soon, they don't care much about cost savings five, ten, or twenty years down the line. They care about costs now, and many times the payment for those downstream procedures will be paid by somebody else, like Medicare. So why prevent osteoporosis if by the time the woman breaks her hip Medicare is responsible for her health

care? We have multiple payers in this country. For that reason no single payer has the incentive to capture all of the potential costs and savings.

Government Agencies

Even government doesn't have all the incentives lined up. Agencies of the government are not the same as The Government. They each have their own annual budgets, and they erect tests of so-called budget neutrality, which means, "We're not allowed to do anything new this year in our Medicaid organization unless it can be offset by some savings." So nothing new that can possibly provide health improvement, maybe even very good value health improvement, will get done unless it is demonstrably cost-saving as well.

And of course none of these decision makers consider patients' or families' time and out-of-pocket expenses, which are part of the cost of health care.

Conclusion

Medical specialty organizations have begun to at least refer to cost-effectiveness analysis in support of practice guidelines. The most recent cholesterol guidelines cite cost-effectiveness analyses showing that it's very cost-effective to give cholesterol-lowering drugs to people after a heart attack, and, as I mentioned, the latest cervical cancer guidelines recognize the value of human papilloma virus testing, and possibly less frequent screening intervals.

Physicians are seeing a proliferation of cost-effectiveness analyses in the leading medical journals. In fact, the *Annals of Internal Medicine*, the journal of the American College of Physicians, has a special structured abstract form for people who write cost-effectiveness studies. Those of us who do those studies and publish in that journal gladly adhere to those guidelines. So at least some physicians are reading these studies. They're going out there in the real world as opinion leaders, educators, and practitioners and beginning to make known that as physicians they have a responsibility to be gatekeepers—informed, educated, sensitive to patient needs, but nonetheless gatekeepers.

We can't do everything for everybody. We've got to make choices. But we're in a position to do it in both a humane and knowledgeable way.

Can cost-effectiveness analysis contain health care costs? Maybe it will help. Will cost-effectiveness analysis improve health? I say yes, but only if it is used.

Bibliography

AIDSinfo. 2004. "HIV and Its Treatment: What You Should Know. Health Information for Patients." Rockville, MD: National Institutes of Health. October. http://www.aidsinfo.nih.gov/ other/cbrochure/english/cbrochure_en.pdf.

Association of Reproductive Health Professionals. 2001(electronic version). "Human Papillomavirus and Cervical Cancer." *Clinical Proceedings* (March 30) http://www.arhp.org/ healthcareproviders/onlinepublications/clinicalproceedings/cph pv/contents.cfm?ID=92.

Bell, CM, RH Chapman, PW Stone, EA Sandberg, and PJ Neumann. 2001. "An Off-the-Shelf Help List: A Comprehensive Catalog of Preference Scores from Published Cost-Utility Analyses." *Medical Decision Making* 21 (4) (July-August): 288-294. http://www.hcra.harvard.edu/pdf/preferencescores.pdf.

Bloom, BS. 2004. "Use of Formal Benefit/Cost Evaluations in Health System Decision Making." *American Journal of Managed Care* 10 (5): 329-335.

Centers for Disease Control and Prevention. National Center for Chronic Disease Prevention and Health Promotion. 2001. "2001 BRFSS Summary Prevalence Report." Atlanta, GA: CDC.

Chapman, RH, PW Stone, EA Sandberg, C Bell, and PJ Neumann. 2000. "A Comprehensive League Table of Cost-Utility Ratios and a Sub-Table of 'Panel-Worthy' Studies." *Medical Decision Making* 20 (4) (October-December): 451-467.

Coffield, Ashley B., Michael V. Maciosek, J. Michael McGinnis, Jeffrey R. Harris, Blake Caldwell, Steven M. Teutsch, David Atkins, Jordan H. Richland, and Anne Haddix. 2001. "Priorities among Recommended Clinical Preventive Services." *American Journal of Preventive Medicine* 21 (1) (July): 1-9.

Congressional Budget Office. 2004. "An Analysis of the Literature on Disease Management Programs." Washington, DC: Congressional Budget Office. October 13. http://www.cbo.gov/ftpdocs/59xx/doc5909/10-13-DiseaseMngmnt.pdf.

Cutler, David M., and Elizabeth Richardson. 1998. "The Value of Health: 1970-1990." *American Economic Review* 88 (2) (May): 97-100.

DePalo, G. 2004. "Cervical Precancer and Cancer, Past, Present and Future." *European Journal of Gynaecological Oncology* 25 (3): 269-278.

Edelson Jonathan T., Milton C. Weinstein, Anna N.A. Tosteson, Lawrence Williams, Thomas H. Lee, and Lee Goldman. 1990. "Long-Term Cost-Effectiveness of Various Initial Monotherapies for Mild to Moderate Hypertension." *Journal of the American Medical Association* 263 (3) (January 19): 408-413.

Elkin, Elena B., Milton C. Weinstein, Eric P. Winer, Karen M. Kuntz, Stuart J. Schnitt, and Jane C. Weeks. 2004. "HER-2 Testing and Trastuzumab Therapy for Metastatic Breast Cancer: A Cost-Effectiveness Analysis." *Journal of Clinical Oncology* 22 (5) (March 1): 854-863.

Enthoven, Alain C. 1980. *Health Plan: The Only Practical Solution to the Soaring Cost of Medical Care.* Reading, MA: Addison-Wesley Publishing Company.

Frazier A. Lindsay, Graham A. Colditz, Charles S. Fuchs, and Karen M. Kuntz. 2000. "Cost-Effectiveness of Screening for Colorectal Cancer in the General Population." *Journal of the American Medical Association* 284 (15) (October 18):1954-1961.

Freedberg, Kenneth A., Elene Losina, Milton C. Weinstein, A. David
Paltiel, Calvin J. Cohen, George R. Seage III, Donald E.
Craven, Hong Zhang, April D. Kimmel, and Sue J. Goldie.
2001. "The Cost-Effectiveness of Combination Antiretroviral
Therapy for HIV Disease." *New England Journal of Medicine*
344 (11) (March 15): 824-831.

Goldman, Lee, Milton C. Weinstein, Paula A. Goldman, and Lawrence
W. Williams. 1991. "Cost-Effectiveness of HMG-CoA
Reductase Inhibition for Primary and Secondary Prevention of
Coronary Heart Disease." *Journal of the American Medical
Association* 265 (9) (March 6): 1145-1151.

Gottschalk, Carl W. 1967. *Report of the Committee on Chronic Kidney
Disease.* Washington, DC: Government Printing Office.

Harvard Center for Risk Analysis. 2004. "The CEA Registry:
Standardizing the Methods and Practices of Cost-Effectiveness
Analysis." Cambridge, MA. Last updated on 29 April 2004.
Available from http://www.hsph. harvard.edu/cearegistry/.

kaiserEDU.org. 2004. "Reference Libraries: HIV/AIDS in the United
States: The Basics." Last updated on November 2004.
http://www.kaiseredu.org/ReferenceLibraries/domestichiv/inde
x.cfm.

Kim, Jane J., Thomas C. Wright, and Sue J. Goldie. 2002. "Cost-
Effectiveness of Alternative Triage Strategies for Atypical
Squamous Cells of Undetermined Significance." *Journal of the
American Medical Association* 287 (18) (May 8): 2382-2390.

Klarman, H. E., J. O. Francis, and G. D. Rosenthal. 1968. "Cost
Effectiveness Analysis Applied to the Treatment of Chronic
Renal Disease." *Medical Care* 6 (1): 48-54.

Kuntz Karen M., Joel Tsevat, Lee Goldman, and Milton C. Weinstein.
1996. "Cost-Effectiveness of Routine Coronary Angiography
after Acute Myocardial Infarction." *Circulation* 94 (5)
(September 1): 957-965.

McGlynn, Elizabeth A., Steven M. Asch, John Adams, Joan Keesey,
Jennifer Hicks, Alison DeCristofaro, and Eve A. Kerr. 2003.

"The Quality of Health Care Delivered to Adults in the United States." *New England Journal of Medicine* 348 (26) (June 26): 2635-2645.

National Cancer Institute. 2004. "Human Papillomaviruses and Cancer: Questions and Answers." Last updated on March 3, 2004. http://cis.nci.nih.gov/fact/3_20.htm.

National Committee for Quality Assurance. 2004. "2004 State of Health Care Quality Report." Washington, DC: National Committee for Quality Assurance. http://www.ncqa.org/ communications/SOMC/SOHC2004.pdf .

National Institute of Allergy and Infectious Diseases. 2004. "How HIV Causes AIDS." Bethesda, MD. Last updated on November 19. http://www.niaid.nih.gov/factsheets/howhiv.htm.

Neumann, Peter J. 2004. *Using Cost-Effectiveness Analysis to Improve Health Care: Opportunities and Barriers*. New York: Oxford University Press.

Plante, C. L. 2000. "1971 Medicare Amendment: Reflections on the Passage of the End-Stage Renal Disease Medicare Program." *American Journal of Kidney Diseases* 35 (4) (April): S45-S48.

Prosser, Lisa A., Aaron A. Stinnett, Paula A. Goldman, Lawrence W. Williams, Maria G. M. Hunink, Lee Goldman, and Milton C. Weinstein. 2000. "Cost-Effectiveness of Cholesterol-Lowering Therapies According to Selected Patient Characteristics." *Annals of Internal Medicine* 132 (10) (May 16): 769-779.

Rettig, Richard A. 1981. "Formal Analysis, Policy Formulation, and End-Stage Renal Disease." NTIS order #PB81-209769. Washington, DC: Government Printing Office. April. http://www.wws.princeton.edu/cgi-bin/byteserv.prl/~ota/disk3/ 1981/8114/8114.PDF.

Rettig, Richard A., and Norman G. Levinsky (editors). 1991. *Kidney Failure and the Federal Government.* Washington, DC: National Academy Press. http://www.nap.edu/openbook/ 0309044324/html/index.html.

Saslow, Debbie, Carolyn D. Runowicz, Diane Solomon, Anna-Barbara
Moscicki, Robert A. Smith, Harmon J. Eyre, and Carmel
Cohen. 2002. "American Cancer Society Guideline for the
Early Detection of Cervical Neoplasia and Cancer." *CA: A
Cancer Journal for Clinicians* 52 (6) (November-December):
342-362. http://caonline.amcancersoc.org/cgi/content/full/
52/6/342.

U.S. Congress. Office of Technology Assessment. 1980. "The
Implications of Cost-Effectiveness Analysis of Medical
Technology: Methodological Issues and Literature Review."
NTIS order #PB81-144602. Washington, DC: U.S.
Government Printing Office. September.
http://www.wws.princeton.edu/cgi-bin/byteserv.prl/~ota/
disk3/1980/8013.html.

U.S. Congress. Office of Technology Assessment. 1980. "The
Implications of Cost-Effectiveness Analysis of Medical
Technology." NTIS order #PB80-216864. Washington, DC: U.
S. Government Printing Office. August.
http://www.wws.princeton.edu/cgi-bin/byteserv.prl/~ota/
disk3/1980/8011.html.

U.S. Congress. Office of Technology Assessment. 1996. "The OTA
Legacy: 1972-1995." Accessed October 27, 2004. Available
from http://www.wws.princeton.edu/~ota/.

Weinstein, Milton C., Sue J. Goldie, Elena Losina, Calvin J. Cohen,
John D. Baxter, Hong Zhang, April D. Kimmel, and Kenneth
A. Freedberg. 2001. "Use of Genotypic Resistance Testing to
Guide HIV Therapy: Clinical Impact and Cost-Effectiveness."
Annals of Internal Medicine 134 (6) (March 20): 440-560.

Winkelmayer, Wolfgang C., Milton C. Weinstein, Murray A.
Mittleman, Robert J. Glynn, and Joseph S. Pliskin. 2002.
"Health Economic Evaluations: The Special Case of End-Stage
Renal Disease Treatment." *Medical Decision Making* 22
(September-October): 417-430.