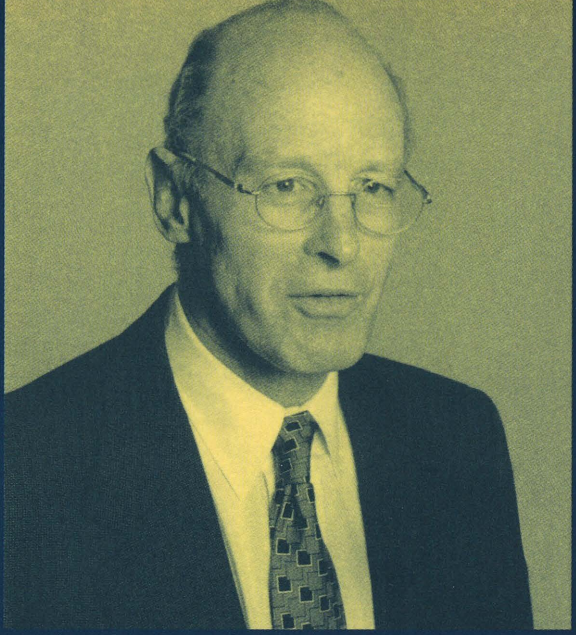

DOCTORS, ECONOMICS AND CLINICAL PRACTICE GUIDELINES:

Can they be brought together?

David M Eddy

OHE
Annual Lecture 1999

DOCTORS, ECONOMICS AND CLINICAL PRACTICE GUIDELINES:

Can they be brought together?

David M Eddy

Annual Lecture 1999

The Office of Health Economics
12 Whitehall London SW1A 2DY
Telephone: +44 (0)20 7930 9203

DAVID EDDY

David Eddy is a Senior Advisor for Kaiser Permanente, the largest not-for-profit Health Maintenance Organisation in the USA, and a longstanding columnist for the Journal of the American Medical Association. A collection of his essays for JAMA on the ethics and economics of clinical practice guidelines has recently been published entitled 'Clinical Decision Making from Theory to Practice'. The material is based on his practical experience of, amongst other things, helping to assess technologies and devise clinical practice guidelines for Kaiser and helping to persuade doctors of the importance of following them.

After qualifying as a doctor, David Eddy took a PhD in Engineering Economic Systems at Stanford University, becoming a Professor of Engineering and Medicine at Stanford. He then became the J. Alexander McMahon Professor of Health Policy Research and Education at Duke University before leaving to become an independent researcher and Senior Advisor to Kaiser. He has held a number of advisory positions for WHO and for the Institute for Medicine and was a White House employee and member of Hilary Rodham Clinton's Health Task Force in 1993.

Office of Health Economics

The Office of Health Economics was founded in 1962. Its terms of reference are to:

- commission and undertake research on the economics of health and health care;
- collect and analyse health and health care data from the UK and other countries;
- disseminate the results of this work and stimulate discussion of them and their policy implications.

The Office of Health Economics is supported by an annual grant from the Association of the British Pharmaceutical Industry and by sales of publications, and welcomes financial support from other bodies interested in its work.

For recent OHE publications see inside back cover.

CONTENTS

DOCTORS, ECONOMICS AND CLINICAL PRACTICE GUIDELINES: Can they be brought together?

The topic I have been asked to address, 'Doctors, Economics and Clinical Practice Guidelines: Can they be Brought Together', is both difficult and controversial. It is also timely. With the creation of the National Institute for Clinical Excellence (NICE), the National Service Frameworks, and the clinical governance project, the UK has an opportunity to develop systematic national solutions to problems that have challenged every society — how to balance the quality and cost of health care in a way that respects both people's humanity and their pocketbooks. My objective in this talk is to contribute in a small way to that effort.

I will begin by discussing the relationship between doctors and guidelines. Then I will address the role of economics in the design of guidelines. Finally, I will try to connect doctors to economics.

I DOCTORS AND GUIDELINES

For this talk I will define a guideline as a simple, operational recommendation about a preferred practice. A statement that women should be screened with Pap smears at least every three years is an example. The hallmarks of a good guideline are simplicity and clarity. It does not try to explain why the practice is being recommended, the factors that were considered in its design, or what the possible outcomes might be; it just gives the answer.

The relationship between doctors and guidelines can be characterized as 'love-hate'. The love part arises from the fact that guidelines are an absolutely necessary part of the practice of medicine. The fundamental problem that makes guidelines necessary is that the appropriate practice of medicine is far too complex for the unaided human mind. In addition to the enormous complexity of biology and disease, and the huge number of variables that affect patients' outcomes, there is the difficulty of identifying and interpreting all the evidence about all the treatments. (I will use the word 'treatment' in a general sense to include all types of health interventions.) If every physician had to work through every clinical decision on his or her

own — for each possible treatment: identifying all the possible outcomes, estimating their probabilities or magnitudes, taking into account every patient's individual characteristics, and weighing the good outcomes against the bad according to one's perceptions of each patient's preferences — we would have either paralysis or chaos. For example, how many physicians who recommend Pap smears really know the probabilities a woman will get or die of cervical cancer; as a function of her age, other risk factors, and screening history; with and without screening; if it is done every one-, two-, three-, four-, or five-years? Guidelines help solve this problem by creating relatively simple 'If…., then…' statements that apply to clinically similar groups of patients. They reduce the physician's task from having to solve an extremely complex cognitive problem down to knowing the guidelines and ensuring that the right ones are applied correctly to the right patients.[1] Accomplishing this still requires great skill and compassion, but it is far more manageable than trying to think through every decision from scratch.

If I am correct that guidelines are an indispensable part of medical decision-making, we would expect them to be common. Indeed, they are. They have been used for centuries if not millennia. They have been found in Egyptian papyruses. 'Eye of newt, toe of frog' was a guideline. It was an incorrect guideline; but it was a guideline. Today they are everywhere. Open any textbook and begin reading; you will find guideline after guideline. Textbooks don't say 'Here are 50 factors and numbers for you to consider'. They say 'Patients with [these characteristics] should be treated with [these treatments]' or 'The indications for [some diagnostic test] are…'. These are guidelines.

If guidelines are necessary, time-honored, and ubiquitous, we then have to ask why they are so controversial — what is the 'hate' part of the love-hate relationship? A commonly given reason is that many physicians do not like to think that they are following guidelines. They prefer to think that they are reasoning through all the pros and cons of every decision from scratch, tailoring their knowledge and insights to every patient's individual characteristics. The complaint about 'cookbook medicine' addresses the sentiment. But I do not believe that this is the main reason because, until quite recently, physicians have happily accepted the guidelines that appear in their textbooks and journals, and those passed on to them from their chief residents and specialty societies. I believe physicians' dislike of

guidelines is relatively new and derives not from the fact that they are guidelines per se, but from recent changes in how guidelines are being developed and in the factors being taken into account.

In the past, guidelines were developed solely by physicians themselves through a very informal, implicit process, based largely on the subjective judgments of practitioners and testimony of experts, and on observations of common practices (e.g. 'standard and accepted practices', 'community standards'). Given the lack of a systematic process or methodological rigor, it might be more accurate to say that guidelines 'evolved', rather than that they were 'developed'.[2] Today, all of that is changing. Guidelines are no longer evolving in a friendly, physician-controlled environment; they are being designed by formally constituted committees using carefully defined methods. What will be happening at NICE and the other centres in the UK that are designing guidelines is very different from the traditional, internally controlled process with which physicians have become comfortable. The reason this tends to be controversial is that, from the physician's point of view, it decreases their control over guidelines and therefore over their practices. A process that used to be entirely under their own control, now has to be shared or negotiated with 'outsiders' like statisticians, epidemiologists and economists from academic groups and government agencies. This in turn introduces elements of uncertainty, distrust and resentment.

In this movement toward a more formal and rigorous process, there are two main issues that are especially difficult for doctors to accept. One has to do with the source of the guideline: should they be derived primarily from their professional beliefs — as expressed through expert opinion, professional consensus and observations of common practices — or should they be derived from an objective examination of actual evidence, even if the evidence conflicts with professional opinion? The second issue has to do with the role of costs. Should the financial cost of a treatment be included in the design of a guideline and, if so, how?

Both of these issues are extremely important and difficult. In this talk I will focus on the second. About the first, I will only say here that I believe guidelines should be anchored to real evidence, not professional judgement, even if there is a broad consensus.[3,4,5] Operationally, I believe we should not design affirmative guidelines — guidelines that promote the use of a treatment — unless there is good evidence from well-designed studies that it is effective and beneficial

in improving health outcomes (that its benefits outweigh its harms). I take this position because of decades of research showing wide variations in practices and high rates of inappropriate care — care that does not match what we know to be true from the evidence, and because of innumerable examples of adamant professional opinion that turned out to be wrong. For the remainder of this paper I will be referring to treatments for which there is good evidence of effectiveness.

Concerning the role of costs in the design of guidelines, I will try to answer two sets of questions. First, is it appropriate to consider costs in the design of guidelines, or should we base guidelines strictly on whether a treatment is effective and beneficial (its benefits outweigh its harms)? Asked another way, isn't that rationing? And isn't that bad? The second set of questions arises if we can agree that costs should be considered in the design of guidelines. If it is ethical to consider costs, how should we do it? In particular, how should we determine whether a treatment is 'cost-effective', and whose values should we use to make that determination?

II ECONOMICS AND GUIDELINES

So that we do not lose the forest for the trees, I will begin with brief answers to these questions. For the first, I will argue that for the modern practice of medicine it is not only ethical but necessary to consider costs when determining the appropriate use of a treatment. I will also argue that this is rationing, but it is good, not bad to ration. With respect to the methods that should be used to include costs in the design of guidelines, I do not believe it is possible today to take a classical economic approach that systematically ranks all treatments by their cost-effectiveness, and marches down the list until the budget is expended. Nor do I believe that it is possible yet to identify a 'correct' threshold for cost-effectiveness — such as the amount of money we should pay for a quality adjusted life year (QALY). Instead, I believe we will need to develop an approach that enables us to make decisions about treatments one by one, even if we do not know the cost-effectiveness of other treatments or a correct cost-effectiveness threshold.

To accomplish that, I will propose a two-step process. First, we should determine if, in terms of its cost-effectiveness, a treatment is an obvious winner, an obvious loser, or something in between. Obvious winners should be the subjects of affirmative guidelines. Obvious losers should be the subjects of negative guidelines (that recommend against their use). For treatments that are not obviously winners or losers, more work is required. I will propose that when the cost-effectiveness of an important treatment is not obvious, we should describe the treatment's possible outcomes to people who are candidates for the treatment and ask them whether they are willing to pay the costs in order to receive the treatment. That will tell us if the treatment is above their cost-effectiveness threshold. And it is the thresholds of the people who will live and die by the results, and who will one way or another pay the costs, that we ultimately care about.

Having given away the answers, I will now provide my reasoning and some examples.

III IS IT APPROPRIATE TO CONSIDER COSTS IN DECISIONS ABOUT TREATMENTS?

To answer this question all we need to do is acknowledge two facts. First, there is a limit on the National Health Services (NHS) budget. Second, there are more things that we can do, in the sense that they are effective and provide at least some benefit, then we can pay for. Taken together these two facts; imply two more facts: the available opportunities exceed the limits, and choices have to be made. Furthermore, those choices will unavoidably involve some trade-offs between financial costs and health benefits. The fact that these choices imply trade-offs between costs and benefits is not an ethical issue that we need to debate. It is just a fact; something that we have to live with.

Now, there are ethical problems that arise when we try to include costs in the design guidelines, but they are not *whether* costs should be included, they are about *how* they should be included. If there is to be any hope of using the available resources in a way that is fair, equitable, efficient, or optimal, we have to consider the amount of benefit we get from allocating the limited resources to different treatments.[6,7,8] This in turn requires that we consider the costs of the treatments. The alternative of ignoring costs would mean that the choices would have to be based on other factors. What other factors might we use if we do not use the amount of benefit a treatment provides for its cost: effectiveness alone? severity of the outcomes? number of people affected? professional preferences? favoritism? squeaking wheels? Each of these has far more serious flaws. If you are not convinced, try to describe an alternative basis for distributing the limited resources across all the worthy patients and argue that it is fair, efficient and ethical.

IV GETTING PAST THE 'RATIONING TABOO'

This type of talk about costs brings up what is probably the most feared word in medicine — 'rationing'. It is critically important that we get past this fear. The dictionary defines 'rationing' as 'the equitable distribution of limited resources'. What is so bad about that? What would we prefer, an inequitable distribution of limited resources? Of course not. What really scares us about rationing in medicine is that we think of it in a very different way. The common connotation is more like 'an arbitrary decision by an anonymous, heartless bureaucrat not to pay for some treatment that has known benefit for a patient, simply to save money, and without appropriate consideration of the value of the treatment to the patient'.

The common connotation of rationing twists the dictionary definition in at least three important ways. First, it omits the fact that resources are limited. This takes away the idea that the allocation was unavoidable, and replaces it with an idea that there really was no need to restrict access to any treatments. The implication is that the restrictions are motivated by stinginess, meanness, or, in the private sector, greed. Second, the common connotation focuses on the treatment of one individual patient. This omits the crucial idea of rationing as a distribution of resources across patients, and strips away the positive concept of fairness. Third, the vision of an anonymous, heartless bureaucrat inserts implications of arbitrariness, lack of consideration of the individual patient's needs and lack of caring. Redefined in this way, of course we would find the idea of rationing to be abhorrent.

To get past this barrier we need to make clear that the common connotation of rationing is *not* what we are proposing for health care. The limitation on services in the NHS is not motivated by meanness or profits; it is an inevitable consequence of the fact that there is a limit to the amount of taxes people can pay, and those tax revenues have to be distributed across a variety of important social services. Anyone who disagrees with this statement can simply tell their MPs (and their neighbors) that they and their neighbors want to have their taxes increased to cover any increases in health care costs, no matter how high they might go. Another crucial difference between the common understanding of rationing and what we are doing is that the people affected are not just one individual patient from whom a treatment is being withheld, but everyone who receives care from the

NHS, all of whom are equally deserving. A third difference is that the decisions are not arbitrary; they are carefully planned to achieve the most health care possible within the limited budget. (We can make the last claim if the guidelines are based on cost-effectiveness. Otherwise I cannot vouch for the careful planning.)

The term 'priority setting' is often used as a euphemistic alternative to 'rationing'. Given the intensity of the misunderstanding and resistance to the word 'rationing', this may be a reasonable alternative for public use. The term 'priority setting' does capture much better the ideas of limited resources, multiple people, fairness, and careful planning. However, to help us get used to the idea that rationing is a good thing, not bad, I will continue to use that word in this paper.

V RATIONING IN THE UK

It is useful to consider how the UK health system currently rations, if for no other reason than to celebrate its strengths. With some admitted simplification, we can identify four main steps that together create rationing in the UK. The first step occurs in the Parliament, when it determines the three-year, cash-limited budget. Because it formally and explicitly imposes limits on the resources available for health care, and because it specifies a particular budget, this is a huge step in the rationing process. The next step occurs when the Department of Health and the NHS Executive (to take the example of England) allocate that national budget to the health authorities, considering factors designed to make the allocations as fair as possible.

The health authorities then have the third step. They have to allocate their budgets to the various entities that will either allocate the resources further, or deliver care. Those entities include the Primary Care Groups (PCGs), the trusts, public health programs, and so forth. In some cases, the allocation decisions are made down to the level of individual treatments. For example, the head of a health authority might specify that a hospital will be paid for 50 major transplants in the coming year, but no more. General medical practitioners who are part of the budget-holding PCGs have to allocate resources either to care for individual patients or to purchase services from the trusts. Finally, in the last step the trusts make allocation decisions when they choose treatments for their individual patients. Thus, in the UK the rationing occurs in a cascade of steps, with different parts of the rationing decision being handled at different levels of the healthcare system.

Let us think about this a bit more, because I believe it is a successful part of the UK system for rationing. In a simplified way we can think of rationing as consisting of three basic types of decisions. The first is a decision that the budget will be limited. An unstated but understood implication of this is that it will not be possible to pay for everything that might possibly have benefit — there will be some limits on the care that can be delivered. The second set of decisions involves what size the budget will be. As I said above, these decisions occur in a cascade, with the Parliament deciding on the overall budget, the health authorities carving it up further, and so forth. The third category includes the specific decisions by physicians when they say to individual patients, 'Mrs Smith, I am very sorry, but I cannot give you

this transplant'. All three of these decisions are part of rationing, but they play very different roles. The first two are general decisions, in the sense that they are made without any reference to specific patients. Decisions in the third category are specific in the sense that they address particular patients and treatments. In the UK, the first two decisions are made at the top and middle levels of the government, whereas the specific decisions are made at the middle and ground levels, so to speak.[9]

There are two things I especially like about this system. First, your Parliament has explicitly decided to limit the budget. This is the 'societal decision' we all talk about that sets in motion, legitimizes, and makes ethical all the subsequent rationing decisions. Second, by having different decisions made by different groups of people, the anguish and the heat is spread more evenly throughout the system. Politicians and health authorities take heat for the general decisions, but that heat is attenuated by the fact that there are no personal faces or anecdotes attached to those decisions. Once the general budget decisions are made, the jobs of the budget-holding PCGs and trusts become much easier. They do not have to suffer any anxiety or take any heat about *whether* there will be a limit on the health care budget or what the limit will be. They are simply given this information as conditions within which they must do their work. Their concerns then become much narrower: how their fixed budget should be allocated to individual patients. These are still agonizing decisions, but they are a lot easier than having to decide both whether treatments will be limited and precisely how.

Please note: I am not saying that rationing in the UK is easy or popular. I am simply saying that, because the rationing decisions are explicitly sanctioned at the top level of government, and spread throughout the system, they are easier and less unpopular than they would be if all aspects of rationing fell on one party — such as the practising physician.

VI IS RATIONING ETHICAL?

Now let us ask whether those three rationing decisions are ethical. Is it ethical for politicians to decide how to distribute the nation's resources across health, education, housing, defense, parks, and any number of other things, even if it means that some effective treatments will not be done? Of course it is. Indeed, it is their job; it is what we vote them into office to do. It would be a dereliction of duty and, I would say, unethical if they did not make those types of allocation decisions.

Once those decisions have been made, the ethics of the remaining rationing decisions becomes obvious. If a PCG or a trust has a limited budget, is it ethical for them to try to allocate that limited budget in a way that is most fair — that improves the health of all the people they serve to the greatest extent possible, within the limits of their budgets? Again, of course it is. Again, it would be unethical for them to allocate their budgets any other way.

VII HOW DID WE REACH THIS POINT?

Even if rationing makes sense from a technical or cerebral pointed view, it still makes us very uncomfortable. I cannot take that discomfort away. However, it can be helpful to understand the reason why health care is driven to rationing. The need to ration is not imposed by any evil politicians, bureaucrats, insurers, executives of managed-care organizations, medical directors or providers. The need to ration is a natural response to a natural phenomenon. The fundamental problem that forces the need for rationing is that the science of health and disease and the development of treatments are increasing at a rate that causes healthcare costs to grow about two to three times faster than other sectors of the economy. If healthcare services are allowed to increase at their 'natural' pace (as driven by the natural growth of knowledge and technology) then healthcare costs will grow much faster than our ability to pay for them, as determined by people's paychecks and government budgets.

As people begin to realize that they can not continue forever to pay premiums and taxes that grow two to three times as fast as their paychecks, they begin to put pressure on the system to control the costs. In the United States that pressure is applied through the marketplace, as both corporate and individual purchasers reject attempts by insurers and managed-care organizations to increase premiums to cover the increases in their costs driven by new technologies. The pressure exerted on insurers and managed-care organizations is then passed on to practitioners through a wide variety of methods designed to manage their use of treatments. In the UK, the pressure is applied directly through the government, as politicians realize that they cannot continue to increase taxes to cover disproportionately rapid increases in health care costs. In both countries, limited payments into the system necessarily create limits on the resources available for providing care. It is this development that introduces the need for some way to distribute the limited resource equitably, ie rationing.

As I stated above, the limit on resources is not an ethical issue that we need to debate, it is simply a fact of life that defines the conditions under which administrators and providers must deliver care. There *is* an ethical issue that providers must address, but it is not whether there should be a limit on resources, it is how the providers should allocate the limited resources in a way that is fair to everyone.

VIII RATIONING IN THE UNITED STATES

At this point, a comparison between rationing in the United States and the UK might be helpful. Both countries face limited resources because of people's inability to pay premiums that outstrip their paychecks. But there is a crucial difference. In the UK, the government has explicitly decided that there will be a limit, and what that limit will be. This narrows the practitioner's task to determining how the limited budget will be spent. In the United States, the political decision to limit the health care budget has never been made. In 1993, when health care reform was being considered, there was an opportunity for US national leaders to announce a national policy to limit the growth of health care costs. But they were unwilling to do that. Instead, they choose to foster the myth that health care providers should be able to provide unlimited care. But that does not mean that they agreed to the obvious consequence of unlimited care — that health care costs will have to grow at the pace of unlimited care. Indeed, we have just the opposite: although individuals, businesses, and government programs are demanding unlimited care, they are also demanding that cost increases be controlled to the rate of growth of the GDP. The effect of all this is that the full heat of rationing falls on the middle and ground levels: insurers, HMOs, medical groups and individual providers. They have to make all the rationing decisions — *that* resources are limited, *what* limit will be, and *how* the limits will be applied to particular treatments and patients. And they must do this in an environment of entirely unrealistic expectations. If rationing in the UK is uncomfortable, in the US it is totally dysfunctional, with no relief in sight.

IX METHODS FOR TAKING COSTS INTO ACCOUNT IN GUIDELINES

Now let us talk about how to allocate resources in a way that maximizes the benefit delivered to a population while staying within a fixed budget. Ideally, we would determine the amount of benefit each treatment is expected to provide, determine the cost of the treatment, and rank all of the treatments according to the amount of benefit they provide divided by their cost (their 'cost-effectiveness'). Then we would march down the list until the budget was expended. Only treatments that made the cut would be included in affirmative clinical practice guidelines. Others would be the subjects of negative guidelines. That would be the classical economic solution to this problem.

Unfortunately, I do not believe it is possible to take that approach. First, our measures of effectiveness, while useful and improving, are still far short of capturing all the nuances of the benefits that treatments can provide. As far as costs are concerned, while there is an acceptable measure of cost — pounds sterling — there are other problems with estimating the costs of treatments. For one obvious example, costs can vary enormously depending on the setting and on how downstream clinical events are managed. But the worst problem is that there are simply too many different treatments, patient indications, settings and other variations to address. Take something that seems easy: cervical cancer screening. The cost-effectiveness would be different depending upon the age at which screening begins, the age at which screening ends, the frequency of screening, a woman's specific risk factors, the specific screening technique, technologies to enhance the test's accuracy, the management of positive results, how pre-cancerous lesions are treated, and so forth.

If we cannot rank all treatments simultaneously, it would be highly desirable to be able to make decisions about single treatments one at a time. One way to do that would be to identify a particular cost-effectiveness threshold above which we would say that a treatment should be provided, and below which we would say that it should not. (Here we need to acknowledge an inconsistency in how the term 'cost-effective' is used. Mathematically, the term 'cost-effectiveness' refers to the cost of obtaining a specified amount of benefit, such as £20,000 per QALY. By this terminology, lower is better. However, we also talk about the 'cost-effectiveness' of a treatment in a way that

implies that higher is better, as in 'Pap smear screening every three years is more cost-effective than annual Pap smears.' To avoid confusion, I will talk about a treatment's cost-effectiveness, where higher is better.) The problem with trying to compare a treatment's cost-effectiveness against a designated cost-effectiveness threshold is that we do not know what the appropriate threshold should be. An economist would want to calculate it by ranking all the treatments by their cost-effectiveness and finding the point at which the budget runs out. But that is the approach that I have just argued we cannot do, for technical reasons.

We do know a few things about how the cost-effectiveness threshold should not be calculated. It should not be the cost-effectiveness of the least cost-effective thing that we are already doing. In the United States, when we encounter a new treatment that is very cost-ineffective, say $100,000 per QALY, advocates of the treatment like to justify it by pointing out that we are already paying for things that are worse. The argument for this approach is that, because we have 'accepted' a less cost-effective treatment, then anything as good as or better than it should also be considered acceptable. That reasoning would make sense only if, at the time the decisions were being made about those cost-ineffective things, people really did know their effectiveness, they really did know their costs, they really were responsible for paying the costs, and they really did make conscious decisions in each case that the effectiveness was worth the costs. But that is not what happened. The decisions to fund these cost-ineffective treatments were made on clinical and political grounds, with little if any understanding of the actual health benefits, and, usually, by those with no direct responsibility for actually paying the costs.

I do believe that it is possible to state an upper limit for the average cost-effectiveness of treatments. It has to be less than what I will call the 'GDP limit' — a country's per capita gross domestic product. Imagine that there is a disease that kills everyone in the UK on January 1st. Imagine there is a pill that, if taken on December 31st, will cure everyone for one year, at which point its potency will cease and another pill will have to be taken for the next year. No country could possibly pay more for those pills than the amount of money people will generate during the coming year — the gross domestic product. In the United States, this method would put the upper bound for the average cost-effectiveness of treatments at about

$35,000 (£20,000). Even that would be very generous because it would consume the entire GDP, leaving nothing for other needs such as food, shelter, education, transportation, recreation. Notice that the GDP limit is considerably lower than the reference threshold that is being used more and more in the United States (with little theoretical justification) — of about $50,000 per QALY.

X A PRACTICAL WAY TO DETERMINE IF A TREATMENT IS 'COST-EFFECTIVE'

Fortunately, even if it is not possible to know the 'correct' cost-effectiveness threshold, it is still possible to make decisions about individual treatments one at a time. The method consists of two main steps. The first is to determine if the treatment is a clear winner or loser. Begin by sketching the treatment's cost-effectiveness. I say 'sketch' because we are not trying to determine the precise cost-effectiveness of the treatment, we are only trying to determine if it is a clear winner or loser. For that purpose, the additional accuracy of a full-bore cost-effectiveness analysis would be unnecessary. Research by Miranda Mugford[10] at the University of East Anglia reinforces the idea of using quick and dirty analyses for purposes like this. If a rough estimate of a treatment's cost-effectiveness indicates that it is highly cost-effective, then we should promote the treatment through affirmative guidelines. Notice that for this approach we do not have to know either the treatment's exact cost-effectiveness, or what the 'correct' cost-effectiveness threshold is; we only need to know that a treatment is highly likely to be more cost-effective than any reasonable threshold. On the other hand, if the treatment's cost-effectiveness is very low, then we would discourage its use with a negative guideline.

Here is an example of a clear winner. When a physician takes time with a smoker and personally advises her or him to stop smoking, about 5 per cent of smokers stop, for good. A conservative estimate (i.e. overestimate) of the cost of the physician's time might be \$45(£30). As for the benefit, we know that smoking shortens a person's life expectancy by about eight years, and that people who stop smoking regain their 'lost' life expectancy fairly rapidly. For this sketch we will ignore all the other morbidities and costs associated with smoking, which, if included, would make the case even stronger. With these very crude assumptions, the cost-effectiveness of physician advice to quit smoking is very high; for £75 we can buy a year of life expectancy (£30/(.05 x 8 years)=£75 per life year).

To make the point about quick and dirty analyses, we can vary the assumptions over a wide range and see that the conclusion that this treatment is a winner would still hold. For example, let us say that the quit rate is only 2 per cent, that a physician's time costs £60, and that quitting smoking adds only four years of life expectancy. Even if the original estimate were off by a factor of 10, this treatment would still

be a very good buy at £750 per year of life expectancy. We do not need to know the exact cost-effectiveness of this intervention, or the 'correct' cost-effectiveness threshold, in order to know that it is worthwhile.

Here is an example that does not make the grade. In the United States the Food and Drug Administration has approved alendronate for the prevention of osteoporosis in any postmenopausal woman, even a woman who has a high bone mineral density and a low probability of having an osteoporotic fracture. Now consider the numbers for a 55-year-old, low-risk, postmenopausal woman. (The numbers would be different for an older, higher-risk woman.) An optimistic estimate of the treatment's effectiveness is that taking alendronate would decrease the chance that she would have a hip fracture in the coming year by about 1 in 37,000. (This estimate is optimistic because an effect on fractures has only been shown in very high-risk women. In average- and low-risk women alendronate did not reduce fractures at all.) There might be some long-term benefits from boosting the bone mineral density, but they have not been demonstrated. There also might be some long-term risks because alendronate attaches itself to the bone in a way that could affect bone remodeling. A final factor to consider is that the drug is inconvenient to take and causes gastric distress in about a third of women. In one study the inconvenience and gastric distress combined to keep more than 50 per cent of women from taking the drug as recommended, and about 30 per cent of women discontinued the drug altogether.[11] The annual cost of alendronate is about £300. Ignoring the potential long-term effects and the inconvenience, this represents a cost of about £11 million to prevent one hip fracture. With these numbers, we know enough to say that we would not want to write an affirmative guideline that all 55-year-old low-risk, or even average-risk women should be put on alendronate. We might want to recommend it for high-risk, older women — we should run the numbers to determine that. But when we push the use towards the lower end of the risk spectrum, its lack of cost-effectiveness becomes more and more obvious.

A useful way to appreciate the fact that treatments very widely in their cost-effectiveness, with some being clear winners and others being clear losers, is to consider different things we can do with a fixed budget. For example, in Southern California Kaiser Permanente, which is a health plan covering about 2.4 million people, $10 million

dollars could be used to either: (1) prevent one hip fracture in 55-year-old, low-risk women, if spent on alendronate; or (2) prevent about 340 hip fractures, if spent on hormone replacement therapy; or (3) prevent about 100 breast cancer deaths, if spent on mammography for women 50 to 65; or (4) prevent about 37 sudden deaths, 300 myocardial infarctions and 700 other serious coronary artery events, if spent on cholesterol treatment in very high-risk people; or (5) prevent about 400 cervical cancer deaths, if spent on giving women Pap smears every three years (vs. no screening at all); or (6) prevent five cervical cancer deaths, if spent trying to screen women every year instead of every three years. These estimates include any expected savings that would accrue from preventing bad outcomes such as fractures, heart attacks or invasive cancers. The three points to be made by this example are: that treatments vary enormously in the amount of benefit they provide for their cost; that even a 'single' treatment can vary widely in its cost-effectiveness depending on how it is used (e.g. risk groups, frequency); and that even back of the envelope calculations can often tell us whether a treatment is a clear winner or clear loser.

XI ASKING PEOPLE WHERE TO DRAW THE LINE

The procedure I have just described works well if a treatment happens to be a clear winner or clear loser. It can also be very useful for identifying trades between overused cost-ineffective treatments and underused cost-effective treatments.[12] But it does not tell us what we should do when a treatment is 'in the middle'. When a line is not obvious, how do we draw the line?

My proposal is straightforward: ask people who are candidates for the treatments what they think. Specifically: (1) identify people who are candidates for the treatment, or, if that is not feasible, identify people who represent real candidates as closely as possible; (2) present them with the best available information on the expected benefits and harms of the treatment; (3) present them with information about the costs; (4) create a situation in which they really would have to pay the costs in order to receive the treatment; and (5) see whether they choose to spend their money to receive the treatment or not. If the majority of them do, we can conclude that a majority of the people who would be in a position to benefit from the treatment believe that its benefits justify its costs. That is, for them the treatment's cost-effectiveness is above their personal threshold. Otherwise not. Elsewhere I have called this approach 'rationing by patient choice'.[13]

This proposal contains answers to both whose values should determine if a treatment is cost-effective, and *how* we should determine their values. My proposal for whose values we should use is very simple: it should be the people who will actually live or die according to the treatment, and who, one way or another (e.g. directly through fee-for-service payments, or indirectly through insurance premiums, HMO dues, or taxes) will pay the costs of the treatment.[14] The how is also very simple: try to create as realistic a setting as possible, find a representative sample of candidates for the treatment, and ask them.

This proposal is similar to what economists call the 'willingness-to-pay' method, but with some important differences. 'Willingness-to-pay' usually refers to asking people what they are willing to pay to have or avoid some outcome or health state. For example, a researcher might ask you how much you are willing to pay to avoid having a hip fracture. The answer would then be used to put a value on the outcome, which could then be used in a cost-effectiveness analysis. For that purpose, the willingness to pay methodology has some

serious drawbacks, especially if it is not used with sensitivity to age, gender, race, socioeconomic status and other factors that can affect a person's ability to pay. For a commonly cited example of a misuse, this method might discover that poor people are willing to pay less for some outcomes than rich people. This finding might be used to justify a policy that a treatment has less value for poor people than for rich people, and therefore should be reserved for rich people. (To be fair to the method, despite the frequency with which this complaint is raised I have never seen an example of this type of misuse, and I cannot imagine any competent economist or policy maker using willingness-to-pay in this way.)

However, here I am proposing that the willingness-to-pay methodology be used in a different way. One important difference is that responses would be applied equally to everyone — not segregated to design different guidelines for different people based on their willingness or ability to pay. A second important difference is that the traditional willingness-to-pay method asks a person to state an amount of money that is 'equal' to the outcome. This type of 'analog' question is unnatural and can be very difficult to answer. (For example, how much would you be willing to pay to avoid a hip fracture?) I am proposing that we ask a much simpler question, whether a specified amount of money is more or less valuable than the expected outcomes of the treatment. This type of 'comparative' question is much more natural and easier. (Would you pay £11 million to avoid a hip fracture?) Indeed, this is the type of question we answer for ourselves every time we make any purchase.

But a more important difference is that my whole purpose is different. My proposal is not that we use this method to get measures of benefit for particular outcomes for the purpose of calculating a treatment's cost-effectiveness. My purpose is far more global. It is to try to discover what would happen if there were a truly efficient and properly functioning marketplace in health care. In an efficient marketplace, of the type that Adam Smith would imagine, the question of whether the value of a good or service is worth its cost is determined (in the marketplace) by whether or not people actually choose to spend their money to buy that good or service. If they do, then we can conclude that, for them at least, the value of the good or service outweighs its cost. Conversely, if they do not choose to buy the good or service, we can conclude that for them the value does not outweigh the cost. In a real marketplace, these individual values can

be expressed one by one, with no need for interference by a third party (unless the good is a public good).

Pure fee-for-service (with the patient paying the full cost of the treatment, with no insurance) would accomplish this. But we cannot do that in health care because the cost of many treatments is too high for most people to contemplate, which is why we need some method for sharing financial risks, as insurance does. However we can find ways to simulate a marketplace that reveals people's values, without exposing them to the risk of financial ruin.[15] By looking at a large number of choices made by a large number of people we can draw conclusions about whether, in general, the value of a good or service is worth its cost to people who have an interest in that good or service. With appropriate attention to differences in incomes and other characteristics that may affect people's choices, this method can be used to give us guidance about allocating people's resources to health care treatments. In essence, this proposal attempts to simulate the choices people would make if health care were being delivered in a system of pure fee for service, but without exposing patients to the risks of financial disaster.

A final methodological point concerns the appropriateness of eliciting the values of people who know that they have or will get a particular condition. A basic tenet of distributive justice as articulated by Rawls is that judgments about the distribution of resources should be made by people who are ignorant of what conditions they may get, and therefore will give unbiased weight to all possible conditions and treatments. Here I am appearing to propose just the opposite — I am going directly to people who already have a condition. The justification is that I am asking them a very different type of question. I am not asking them whether public resources (somebody *else's* money) should be spent on a treatment for their condition (we already know they would all say 'yes'); I am asking them whether they are willing to spend their *own* money on a treatment. And my purpose is different. It is not to give them control over the use of public money; it is to learn how people who have a real interest in a treatment compare its benefits to its costs. For that purpose they are the right people to ask.

These ideas can be made more clear with a few examples. Consider the decision whether to use streptokinase or tissue plasminogen activator (tPA) for the acute treatment of heart attacks. tPA is slightly more effective in preventing deaths, but is much more expensive.

How should we decide if its slightly greater benefit is worth its additional cost? We can ask people who are candidates for having heart attacks. For example we might identify a group of people who have angina and other risk factors that put them at a high-risk of having a heart attack, and ask them something like this: 'If you have a heart attack in the coming year, there are two drugs that we could give you to thin your blood and increase your chances of survival. One is called 'tPA'. It would increase your chances of surviving the heart attack by about 0.4 per cent (that is, about 4 in 1,000 or 1 in 250), as compared to the other drug, which is called 'streptokinase'. But tPA will cost you about £1,400 more than streptokinase. If you want to receive the tPA, you will have to pay the full £1,400 additional cost out of your own pocket. That is, the NHS (or your insurance) covers streptokinase but not tPA. Are you willing to pay £1,400 to get tPA instead of streptokinase if you should have a heart attack?'

It is important to understand that there is no right or wrong answer to this question. It simply comes down to how real people weigh improvements in survival probabilities against cost. Let us try it right now. About a quarter of the men reading this paper will have a heart attack at some time in the future and will either get tPA or streptokinase. Ask yourself: would you be willing to pay £1,400 in order to get tPA, which will decrease your chance of dying of a heart attack by 0.4 per cent? Those who say 'no' are telling us that they value the money more than they value what is to them a small reduction in mortality. For them it is not cost-effective. Those who say 'yes' are saying that the increase in chance of survival is large enough to be worth the money. For them it is cost-effective. When I ask this question of large audiences, the great majority say 'no'. If a more formal survey showed the same result, I believe we should honor their choices by having a guideline that says 'For patients with acute myocardial infarctions, give streptokinase not tPA.' Conversely, if a formal survey revealed that most people say 'yes' (and back it up by truly paying the £1,400), I believe we should honor that result and write a clinical guideline to use tPA. By the way, everyone who says 'no' has explicitly said that costs do count, and there is a limit on the amount of money he or she will pay to add more quality — even for a life-or-death outcome.

Let us try another example that applies more to women: mammography screening for women between the ages of 40 and 50.

The cost is about £500 for 10 years of screening (assuming it is done annually). The chance that a 40-year-old woman will get a breast cancer in this age decade is about 1.2 per cent. Without screening, the chance she will die of such a cancer is about 0.8 per cent. Ten years of screening with mammography will not change the chance a woman gets breast-cancer but research has shown that it will reduce the chance that she will die of breast cancer from 0.8 per cent to 0.7 per cent, that is a decrease of about 0.1 per cent, or 1 in 1,000. Stated another way, we would have to screen 1,000 women for 10 years between the ages of 40 and 50 in order to prevent one woman from dying of breast cancer. Another fact a woman will need to know it is that there is 10 per cent to 30 per cent chance of having a false positive mammogram that will require a biopsy. How do we determine if this degree of benefit is worth the risks and cost? Ask 40-year-old women. Are they are willing to pay £500 to get the mammograms? Incidentally, when I ask this question of women in this age group, most say 'no'.

These types of direct questions can be made realistic if the cost of the treatment is within the financial capabilities of the people being asked. They do not work for treatments whose cost are too high to be realistically paid. For example, it would be meaningless to ask most people 'Are you willing to pay £50,000 to receive high dose chemotherapy and bone marrow transplantation for stage IV breast cancer?' because most people do not have £50,000. But we can still ask questions that will provide useful information about how people weigh benefits against costs. The strategy is to identify people who do not yet have, but *might* get a condition that would benefit from a procedure, tell them the probability they will get the condition and therefore become candidates for the treatment, and ask them if they are willing to pay the *expected* cost of the treatment in order to be able to receive it *if* they should get the condition. This is analogous to asking them if they are willing to pay for a rider to an insurance policy that would cover the treatment if they happened to get the disease. To stay with the example of high dose chemotherapy and bone marrow transplantation for stage IV breast cancer, we can explain to healthy women that there is about a 10 per cent chance that they will get breast cancer sometime in their lives and about a 4 per cent chance of developing stage IV disease. The randomized controlled trials completed thus far do not show that the treatment increases survival, although it is still possible that some regimens might provide some

benefit for some patients. An optimistic guess is that high-dose chemotherapy and bone marrow transplantation might possibly improve the chance a woman with stage IV breast-cancer will survive two years by about 5 per cent. We need to stress that the existing evidence is discouraging and the estimate of an improved two-year survival rate is a guess or hope, not a fact. High dose chemotherapy and bone marrow transplantation cost about £50,000. Because a woman has only a 4 per cent chance of ever needing it, the expected cost to her (in an actuarial sense) is about £2,000 (i.e. £50,000 x 0.04). Therefore the question to a woman would be: 'Are you willing to pay £2,000 now, in order to have access to high-dose chemotherapy and bone marrow transplantation if should you ever develop stage IV breast cancer?'

There is another, bolder way to learn how people value benefits vs. costs for very expensive treatments. We can take people who actually have the condition and are currently candidates for the treatment. We then give them an amount of money equal to the cost of the treatment and assure them that the money is really theirs. They can keep it, give it to their heirs, or spend it, as they see fit. We then tell them about the treatment and ask them if they are willing to buy it. We know that they have sufficient money to pay for the treatment without causing them any extra financial hardship, because we have just given them the money. For example, we could give £50,000 to women who actually have stage IV breast cancer, let them savor it in their bank accounts for a while (to make sure they really believe it is theirs to do with as they please), and then ask them if they are willing to spend it to receive high-dose chemotherapy. The answers to these questions would tell us a lot about how people actually compare health outcomes and costs.

Applying this method properly would require considerable attention to such things as: (1) getting as accurate as possible information about the treatment's benefits, harms and costs; (2) including indirect costs and savings in the net costs presented to people; (3) possibly adjusting the costs to account for differences in people's incomes, depending on how the answers are to be used; (4) framing the questions in as unbiased a fashion as possible; (5) helping people understand the nature of the possible health outcomes; (6) selecting a representative sample of people; (7) making the choices as realistic as possible (e.g. ideally the people really would have to pay real money); and (8) aggregating the answers to develop a policy; and so forth.

We would also have to accept the fact that our information will be imperfect, that many people will misunderstand the numbers or the nature of the outcomes, that the effects of framing can never be completely eliminated, and so forth. However, these issues affect every decision in every aspect of our lives. The question is not whether this approach is perfect. It is not. Nor is our system for electing public officials or choosing which automobile to buy or job to take. The question is whether there is a better method. I believe that, because it is real people who will actually suffer the outcomes and eventually pay the costs of health care, we ought to try to get as close to their values as possible. This is a proposal for doing that.

One final comment is irresistible. Asking these types of questions to a broad range of people about a broad range of treatments will quickly reveal the 'correct' cost-effectiveness threshold — the threshold at which the people whose values count the most balance cost and quality. It is this threshold that should determine the overall budget for health care, not the other way around.

XII DOCTORS AND ECONOMICS

If we can agree that doctors need guidelines, and that it is both ethical and feasible to include costs and cost-effectiveness in the design guidelines, the next thing we have to do is help doctors feel comfortable with guidelines that include costs. For example, they should feel comfortable when they apply a guideline for treating heart attack patients with streptokinase instead of tPA, even if they know that tPA is actually superior to streptokinase and is being rejected solely because someone has determined that it costs too much.

The best way to approach this issue is to imagine the alternative — imagine guidelines that are designed strictly on the basis of the degree of effectiveness or some other factor, ignoring costs. Such an approach would say that heart attack patients should be treated with tPA, not streptokinase. Such an approach would also say that average- and low-risk women should be treated with alendronate if they don't mind the inconvenience. Those recommendations would certainly fit what many physicians might consider their duty or 'covenant' with the patient. And that would be fine, except for one overwhelming problem: if we wrote all our guidelines on the basis of effectiveness, ignoring cost, collectively they would break the bank. Unfortunately, taking costs out of guidelines and talking about duties will not increase the budget. So if a guideline does not incorporate cost-effectiveness, then there will be a conflict between what the guideline says physicians should do and what physicians *can* do.

This conflict will be focused directly on physicians. They will be put in the intolerable position of having to make the rationing decisions themselves, choosing which guidelines to follow and which to overturn because they are to expensive. Given that costs will have to be addressed one way or another, for doctors it is far preferable to have the cost-effectiveness trade-offs included in the guidelines, than it is to force them to perform their own personal cost-effectiveness analysis every time they try to apply a guideline to particular patient. The key to accepting the incorporation of costs and cost-effectiveness in guidelines is to recognize that the budget is limited. Incorporating costs in guidelines so that physicians can simply follow the guidelines, is far better than forcing physicians to incorporate cost on their own, in contradiction to the guidelines.

XIII SUMMARY AND CONCLUSIONS

Guidelines are an essential part of medical decision-making. They have been used for centuries. However, increasing pressures to control costs have caused us to change the way guidelines are designed, and the factors they need to include. Of particular importance is the need to incorporate costs and cost-effectiveness in recommendations about the use of treatments. The fundamental problem is that the 'natural' growth of medical knowledge and technology is causing health care costs to grow at approximately two to three times the rate of growth of the general economy, from which we get our paychecks. As a consequence, there is a limit to how much we can pay for health care over the long run. This limit on premiums or taxes in turn creates a limit on the resources available for providing care. In order to maximize the health of people we serve, we need to set priorities according to the amount of benefit a treatment provides, for the resources it uses.

Determining whether a treatment is 'cost-effective' in this sense is very difficult. For methodological and technical reasons, it is not possible to use a classical economic approach of calculating the cost-effectiveness of every treatment and marching down a ranked list until the resources are exhausted. Instead, we need an approach that can determine the appropriateness of treatments one by one. One approach is to perform a rough cost-effectiveness analysis to determine if a treatment is a clear winner or loser, design affirmative guidelines for the former, and negative guidelines for the latter. For treatments that are 'in the middle', I propose that we present information on the treatment's benefits, harms and costs to people who are candidates for the treatment and see whether they believe the treatment's benefits outweigh its costs. This approach will give us the best understanding of how the people who count — those who will actually live and die by the health outcomes, and who one way or another will pay the bills — balance cost and quality. A comforting by-product of this approach is that if people are truly willing to pay more money for higher quality, then this approach will show that. Similarly, if people believe that for some treatments the amount of benefit they provide is not worth their cost, then people will be telling us this directly, which will make the design of guidelines to withhold such treatments far more understandable and defensible.

REFERENCES

1 Eddy DM. 'Clinical policies and the quality of clinical practice.' *New England Journal of Medicine* 1982; 307: 343-347.

2 Eddy DM. 'Practice policies: where do they come from?' *Journal of the American Medical Association* 1990; 263:1265, 1269, 1272, 1275.

3 Eddy DM. 'Practice policies: guidelines for methods.' *Journal of the American Medical Association* 1990; 263:1839-1841.

4 Eddy DM. 'Guidelines for policy statements.' *Journal of the American Medical Association* 1990; 263:2239-2243.

5 Eddy DM. 'Investigational treatments. How strict should we be?' *Journal of the American Medical Association* 1997; 278: 179-185.

6 Eddy DM. 'Cost-effectiveness analysis: will it be accepted?' *Journal of the American Medical Association* 1992; 268:132-136.

7 Eddy DM. 'Cost-effectiveness analysis: is it up to the task?' *Journal of the American Medical Association* 1992; 267:3342-3348.

8 Eddy DM. 'Cost-effectiveness analysis: a conversation with my father.' *Journal of the American Medical Association* 1992; 267:1669-1672, 1674-1675.

9 Eddy DM. 'Broadening the responsibilities of practitioners. The team approach.' *Journal of the American Medical Association* 1993; 269:1849-1855.

10 Mugford M. 'How does the method of cost estimation affect the assessment of cost-effectiveness in health care?' Oxford: DPhil thesis, University of Oxford, 1996.

11 Ettinger B, Pressman A, Schein J, et al. 'Alendronate use among 812 women: prevalence of gastrointestinal complaints, noncompliance with patient instructions, and discontinuation.' *Journal of Managed Care Pharmacy* 4: 488-492.

12 Eddy DM. 'Rationing resources while improving quality: how to get more for less.' *Journal of the American Medical Association* 1994; 272: 817-328.

13 Eddy DM. 'Rationing by patient choice.' *Journal of the American Medical Association* 1991; 265:105-108.

14 Eddy DM. 'What do we do about costs?' *Journal of the American Medical Association* 1990; 264:1161, 1165, 1169-1170.

15 Eddy DM. 'Connecting value and cost: whom do we ask and what do we ask them?' *Journal of the American Medical Association* 1990; 264:1737-1739.