COMMENTARY

What Is the Price of Life and Why Doesn't It Increase at the Rate of Inflation?

AN ARTICLE in the February 2000 issue of *JAMA* concluded that annual retinal screening for many individuals with type 2 diabetes mellitus may not be warranted on grounds of cost-effectiveness. Vijan et al[1] reported that, compared with biannual screening, annual retinopathy screening for low-risk patients with diabetes costs more than $100 000 for each additional quality-adjusted life year (QALY) gained. The results of a study published in the March 2000 issue of the *New England Journal of Medicine* concluded that extending hospital stays beyond 4 days for patients with uncomplicated myocardial infarctions was economically unattractive, costing more than $105 000 per QALY gained.[2] These studies demonstrate that commonly used interventions may not be worthwhile investments of health care resources. By contrast, a study published in the June 2000 issue of the *Annals of Internal Medicine* concluded that, compared with no treatment, sildenafil (Viagra) is a cost-effective treatment for erectile dysfunction, producing an incremental QALY for the relatively low cost of $11 000.[3] The latter study raises questions about whether many health care insurers were hasty in deciding that they would not add sildenafil to the list of services covered by their health plans.[4]

In each of the 3 articles noted above, the authors made prominent mention of a threshold by which cost-effectiveness experts judge the affordability of health care services: those interventions that produce a QALY for $50 000 or less are a bargain, whereas those that require $100 000 or more are considered unaffordable. (A QALY is a measure that allows comparisons between health care interventions that save lives and those that improve quality of life. For example, an intervention that cures a patient of a health condition with a quality of life halfway between perfect health and death yields 0.5 QALYs per year, whereas an intervention that saves a person's life and returns him or her to perfect health yields 1 QALY per year.) The $50 000 to $100 000 QALY cutoff goes back to at least 1982.[5] A commonly used justification for this threshold figure has been that $50 000 per QALY gained was approximately the cost-effectiveness ratio calculated for the use of dialysis for patients with chronic renal failure. Renal dialysis is a federal entitlement to all US citizens under Medicare. As the argument goes, if the US government thinks that dialysis should be offered to all who need it, then interventions with similar or better cost-effectiveness should likewise be offered to everyone.

The $50 000 to $100 000 per QALY gained amount is often cited in the literature as the cost-effectiveness threshold, as reflected in the 3 economic analyses described above. For example, in a review of 228 cost-effectiveness analyses of pharmaceutical interventions, Neumann et al[6] found that the investigators in 34% of the studies refer to the $50 000 per QALY figure when discussing the implications of their findings. Most of the others cite no threshold. Many physicians have expressed skepticism about the role of cost-effectiveness information in making health care decisions.[7,8] Is it any wonder that they are skeptical, given that the $50 000 to $100 000 per QALY threshold has persisted for 2 decades without adjustment? After all, if society was willing to spend $50 000 for a QALY 20 years ago, shouldn't it be willing to spend more for one now?

In this article, we discuss one factor that may contribute to physicians' skepticism about using cost-effectiveness analysis in health resource allocation decisions: that the commonly cited threshold figure of $50 000 to $100 000 per QALY is too low. We then explain that a significantly higher threshold is more consistent with societal willingness to pay for medical interventions. Finally, we argue that the accepted threshold is dynamic and must change over time. Changes in the optimal price of a QALY are dependent not only on inflation but also on complex interactions between social desires to control health care costs and the rate of development of new health care technologies. We conclude that if society hopes to constrain health care cost growth, the inflation-adjusted amount it should spend to produce an additional QALY will likely need to decrease over time if technological innovation continues at its current pace.

THE ARGUMENT FOR A COST-EFFECTIVENESS THRESHOLD

Before determining an appropriate cost-effectiveness threshold (ie, the amount of money that *should* be spent to produce a QALY), it is worth asking whether the idea of having *any* threshold is worthwhile. The appeal of a precise threshold is easy to understand. The exis-

tence and acceptance of such a benchmark would greatly facilitate decision making by reducing the bottom line of any cost-effectiveness study to whether or not the ratio falls above or below the predetermined cutoff value.

Determining the appropriate threshold is a complex task that is dependent on multiple interrelated factors. Decision-makers' willingness to pay for a QALY depends on the context and perspective of the decision itself. A single threshold may not be appropriate for all individuals. Moreover, broad societal resource allocation decisions, such as how much to spend on health care vs education vs a tax cut, may yield one threshold, while narrower decisions, such as how to maximize health benefits within a fixed budget (eg, Medicaid), may yield another. Given the potential for multiple contexts, it is unlikely that a unique, broadly applicable, threshold value per QALY gained exists.

In light of the reality that a single acceptable threshold will not emerge, the US Public Health Service panel on cost-effectiveness concluded that any threshold should be recognized as a *guide* to health care spending decisions, not as a *determinant* of such decisions.[9] In order to guide such decisions, it is useful to be able to refer to some cost-effectiveness threshold, so that decision makers can place new cost-effectiveness analyses into a broader context. Thus, for the purposes of this article, we will assume, along with much of the cost-effectiveness community, that such a benchmark value exists.

EVIDENCE THAT THE $50 000 TO $100 000 THRESHOLD IS TOO LOW

If the desire for a cost-effectiveness threshold exists, the next question to ask is whether the long-cited benchmark of $50 000 to $100 000 per QALY is consistent with societal preferences. Two types of evidence inform this question. The first type is qualitative evidence regarding Americans' level of comfort in basing resource allocation decisions on the existing thresholds. The second type is quantitative evidence based on the results of studies evaluating the statistical value of life. These studies seek to infer people's willingness to pay for years of life or QALYs from observations of actual behavior (eg, wage differentials for riskier occupations) or surveys explicitly incorporating risk/money tradeoffs.

Qualitatively, many members of the medical profession are uncomfortable with the notion of denying health care interventions that have been proved to improve clinical outcomes, even when economic evaluations find their use not to be cost-effective relative to conventional thresholds.[7,10] This discomfort may reflect a general distaste for thresholds at any level (the "save a life at any cost" mentality), which says little about the merits of existing thresholds.[11] Such distaste may arise from distrust of the motives of the proponents of cost-effectiveness studies (eg, insurers or pharmaceutical manufacturers) or from a belief that cost-effectiveness methods are not well defined or consistently applied. Alternatively, this discomfort may not reflect the acceptance of the concept of a threshold, but a perception that conventional thresholds are too low. We suspect that all these factors are contributing to clinicians' discomfort with cost-effectiveness analyses.

The common practice of performing annual Pap smear screening for women at low risk of cervical cancer, despite analyses that indicate cost-effectiveness ratios of greater than $700 000 per QALY gained, probably illustrates discomfort with any threshold.[12] However, public controversy surrounding screening mammography for women aged 40 to 49 years may reflect a willingness to accept a threshold, but one greater than the commonly cited $100 000 per QALY gained. In that case, a National Institutes of Health Consensus Development Panel issued a statement that for this population current data do not support universal screening, which has a cost of approximately $150 000 per QALY gained.[11] The negative reaction to this statement included dissent from 2 of the National Institutes of Health panel's 12 members, a 17-1 vote by the National Cancer Institute's Advisory Board to support screening in this age group, a 98-0 vote by the US Senate in favor of a nonbinding resolution to support screening in this age group, and considerable media coverage.[11] This reaction was dismissed by some as an inappropriate substitution of special interest politics for scientific evidence.[13] Although the politics of the particular disease undoubtedly influenced the reaction, an unwillingness to deem $150 000 per QALY gained as an unworthy investment may also have played a role. Examples from recent coverage decisions for diagnostic tests that improve detection of disease at increased costs (eg, colonoscopy for colon cancer and the *addition* of digital mammography to routine mammography to detect breast cancer) also support the argument that a substantially higher threshold is warranted.

One important caveat to this sort of qualitative evidence concerns the role of the health insurer in the provision of a medical intervention. Since most of the costs of medical care are not borne by patients or providers, insurance coverage creates an incentive to deliver and utilize care that would not provide sufficient expected benefits to justify its costs, if those costs were actually borne by patients or providers. In fact, third-party payment is a large part of the rationale for applying a cost-effectiveness threshold in the first place. Just as we may consider a Mercedes-Benz to be a cost-ineffective alternative to a Chevrolet when spending our own money, how many of us would head to the Chevrolet dealership if someone else offered to buy us any automobile of our choosing?

Other evidence that the $50 000 to $100 000 per QALY threshold is too low is derived from the statistical value-of-life literature. We previously reviewed the estimates (generally, dollars per life) in this literature and converted them to dollars per QALY.[5] The reviewed studies included studies of actual behavior (eg, willingness to pay for safety improvements, the extent to which wages in riskier occupations exceed those in lower risk but otherwise similar occupations) as well as con-

tingent valuation studies (surveys about hypothetical tradeoffs between money and risk). Although this literature has considerable variation in methods and study populations and consequently yields variable estimates of dollars per QALY gained, the clear impression is that the value per QALY gained implied by the statistical value-of-life literature exceeds conventional cost-effectiveness thresholds. The median behavioral or contingent valuation study in our review implied a value per incremental QALY of $265 000 in 1997 US dollars. Furthermore, 28 of the 35 value-of-life estimates we reviewed implied a value per QALY in excess of $100 000.

While we do not wish to overstate our confidence in any specific threshold being correct, both the qualitative evidence and the empirical data from the statistical value-of-life literature suggest that the $50 000 and even $100 000 thresholds are likely conservative estimates of the value of a QALY. Furthermore, it does not appear from multiple clinical examples that the frequently cited $50 000 to $100 000 thresholds are generally enforced in practice.

THE THRESHOLD ONCE CHOSEN MUST BE DYNAMIC

It should not surprise anyone that the $50 000 to $100 000 threshold is too low. The original $50 000 threshold was based on the reasoning that if dialysis cost $50 000 per QALY, then anything else that cost $50 000 or less per QALY should be made available to patients. However, by this reasoning, the $50 000 cutoff should have been the *floor of the threshold,* not *the ceiling*. If offering dialysis to patients compelled society to offer other interventions that cost $50 000 per QALY, who is to say that society should not have been compelled 2 decades ago to spend *more* than $50 000 for a QALY?

For the purposes of argument, let us suppose that society agrees to apply $265 000 per QALY as the benchmark for judging the cost-effectiveness of medical interventions. Coming to this decision leads to 2 questions. First, can the United States really afford to call (and provide) care costing nearly 10 times its per capita income ($29 676 in 2000) *cost-effective*? Second, with enforcement of a such a cutoff, will health care costs be affordable over time?

The key to answering the first question lies in the distinction between the *average cost* and the *marginal cost* of a QALY. The cost-effectiveness threshold selected represents the cutoff for the acceptable marginal cost of a QALY (the most society would be willing to pay to produce 1 *additional* QALY), while the average cost of health care represents the amount of non–health care consumption forgone in order to receive health care on average. Looked at in another way, 50% of the population incurred less than 5% of aggregate health care expenditures. It is this relatively low average cost of health care that allows us to potentially "afford" what may seem to be an exorbitant "marginal cost" of that last additional QALY. That is, we would spend close to $265 000 per QALY for very few individuals in any given year.

Ultimately, the issue of affordability depends on how adoption (and enforcement) of a higher threshold value for defining cost-effectiveness would affect current health care spending and alter spending over time. Because currently cited standards are often ignored, one cannot be certain how a more generous benchmark would influence patient and physician behavior.

IMPACT OF THRESHOLD ON LONG-TERM COSTS

For the sake of illustration, suppose the cost per QALY of all interventions for all potential users was calculated, and the United States adopted and enforced a standard of $265 000 per QALY gained. Relative to the current situation, where the $50 000 to $100 000 standards are not enforced the resultant short-run change in aggregate health expenditures that resulted from adopting the new standard could be either positive or negative. Total health care costs might be decreased through reduction in the use of services that cost more than $265 000 to produce a QALY (eg, the routine Pap smear test). However, if the $100 000 threshold was having a stronger influence on clinical behavior than was previously suspected, the relaxation of the standard could encourage the use of additional services that cost less than $265 000 but more than $100 000 per QALY. This increased use would affect the bottom line of third-party payers. Of course, if thresholds at any level are irrelevant to practice patterns, changing the standard would have no effect on spending.

What about the long run? What would happen to health care expenditures if decision makers consistently followed the $265 000 per QALY cutoff (or any cutoff for that matter)? Our best guess is that health care expenditures would continue to increase faster than the general inflation rate. To illustrate: imagine that a new treatment is developed for a disease. Offering the treatment to a population with this disease costs $200 000 and produces 3 additional QALYs in that population, at an incremental cost-effectiveness ratio of $67 000 per QALY. Because this treatment passes the benchmark, it would be offered to the population. Now imagine that another treatment is developed for this disease that, when used *in addition to* the first treatment, increases health care costs another $200 000 but produces another QALY. It too would be offered. Theoretically, a large number of such incrementally beneficial services could be developed, each raising per capita health care costs but producing enough QALYs to fall beneath the $265 000 threshold.

Furthermore, it is important to remember that new interventions can cost *less* than previous interventions and still raise total health care costs, because they increase health care use, the "behavioral response."[14] For example, on a case-by-case basis, laparoscopic cholecystectomy is a cost-saving alternative to traditional cholecystectomy: the extra costs of laparoscopy are more than made up for by decreased hospital stays.[15] However, these per-case savings can be overwhelmed by the increased number of *additional* patients who are now deemed to be suitable for cholecystectomy (eg, those thought either to be at too high of a risk for tra-

ditional cholecystectomy or to have symptoms too mild to warrant open surgery). This example illustrates that maintaining any arbitrarily defined cost-effectiveness threshold over time is unlikely to constrain health care inflation in a substantial way. New interventions are developed every day that produce additional QALYs for patients for less than the current $50 000 to $100 000 threshold. Of course, even more new interventions would fall under the proposed, higher threshold. Only a rare minority of new interventions lead to net cost savings from the budgetary perspective. Most new pharmaceutical products that come to market have a cost-effectiveness that is substantially better than even the current $50 000 to $100 000 threshold for a QALY.[6] Offering these medications to patients will ultimately drive up health care costs.

HOW SHOULD A COST-EFFECTIVENESS THRESHOLD BE USED?

In theory, a societal cost-effectiveness threshold can be determined if (1) all health care interventions can be ranked from most cost-effective to least cost-effective and (2) a fixed health care budget can be determined. The budget would be spent, in order, from the most cost-effective intervention to the least cost-effective. Budget planners would estimate the cost of offering each intervention to the population in question. Moving down the list, budget planners would estimate how many interventions could be offered before the budget was exhausted. At that point on the list, a line would be drawn such that the more cost-effective interventions, above the line, would be funded and the less cost-effective interventions, below the line, would not. The cost-effectiveness of the last intervention funded would effectively act as the cost-effectiveness threshold.

As Oregon discovered in its Medicaid rationing plan, this approach, while theoretically appealing, is practically flawed.[16,17] To begin with, there is no societal consensus about how much money to spend on health care. Thus, it is difficult to know where on the list to draw the line between funded and unfunded interventions. More importantly, good cost-effectiveness data are lacking for many health care interventions, and the cost effectiveness depends not only on the intervention but also on the characteristics of the patients to whom it is applied. Thus, it is impossible to generate a thorough cost-effectiveness list: hence the appeal of an informal cost-effectiveness threshold (a number of dollars per QALY that can be used to *guide* health care decision making). Formulary committees can appeal to such a cost-effectiveness threshold when deciding the merits of new medications. Physicians can look to the threshold when making diagnostic and treatment decisions. Managed care medical directors, insurance company executives, and government administrators can use the threshold to inform their treatment coverage decisions. In none of these cases should cost-effectiveness data, or the admittedly arbitrary cost-effectiveness threshold, determine any final decision. Instead, cost-effectiveness data need to be incorporated with other information before important decisions can be made.

Our concern is that people's willingness to use cost-effectiveness information is limited, in part, by the implausibility of the current $50 000 to $100 000 per QALY threshold. The threshold is counterintuitively low. And anyone who has been reading cost-effectiveness literature over the years recognizes that the threshold has not been reevaluated for decades. Given the implausibility of the threshold, decision makers may minimize their use of *any* cost-effectiveness data. Revising the threshold now to be greater than $100 000 per QALY, and revisiting the threshold over time, may increase the credibility of cost-effectiveness analyses in guiding health care decisions.

CONCLUSIONS

Cost-effectiveness information deserves a role in health care decision making. Efforts spurred on by the US Public Health Service panel on cost-effectiveness to standardize and raise the quality of cost-effectiveness methods are intended to enhance the credibility of the cost-effectiveness literature and to improve comparability across studies. Nevertheless, without some context about when a cost-effectiveness threshold is acceptable, decision makers will have difficulty making use of cost-effectiveness information. We propose that currently cited thresholds are too low, especially the commonly cited $50 000 per QALY gained cutoff. To support this point, we note several commonly used clinical examples with cost-effectiveness ratios that exceed conventional thresholds. In line with these intuitions, we think that it is more appropriate to raise the cost-effectiveness threshold significantly, approaching $200 000 or more per QALY, after appropriate input from all vested stakeholders. Setting a new threshold is best accomplished by consensus process, such as that used by the US Public Health Service when it established standards for the field of cost-effectiveness in the mid 1990s. We think that the next time such a consensus process is convened, experts should seriously consider revising the cost-effectiveness threshold.

Maybe more important than the setting of the threshold, we believe that the cost-effectiveness threshold should be refined regularly to allow for expected changes in cost of goods (inflation), per capita income, burden of disease, innovation in diagnosis and treatment, and patient preferences. For cost-effectiveness analyses to be credible allocation tools for key decision makers, they need to reflect changing circumstances. While inflation and budgets are likely to increase the threshold over time, we believe that continued innovation-producing interventions that meet the benchmark (and resultant demand for them) will require that the cost-effectiveness threshold actually be reduced over time.

We acknowledge that there is no simple way to determine the appropriate price of a QALY and that whatever the price of a QALY should be now, it likely needs to change over time. At the same time, we need to recognize that health care interventions that seem to be a bargain

may not always be affordable from a population perspective owing to the behavioral response. In the long run, as more and more health care interventions continue to be developed and used, what appears to be an affordable QALY today may not be an affordable QALY tomorrow. Much in the way that a family can bankrupt itself if it purchases too many goods, even when those goods are available at bargain prices, those responsible for health care spending must understand the aggregate budgetary impact of providing reimbursement for interventions that produce one extra QALY for what appears to be an affordable price. For now, the price of a QALY should be significantly greater than $50 000 to $100 000. But the price needs to be revisited on a regular basis in order to forge some consensus about what society can continue to afford.

Peter A. Ubel, MD
University of Michigan Health Systems
300 N Ingalls, Room 7C27
Ann Arbor, MI 48109-0429
(e-mail: paubel@umich.edu)
Richard A. Hirth, PhD
Michael E. Chernew, PhD
A. Mark Fendrick, MD
Ann Arbor

Dr Ubel is a recipient of a career development award in health services research from the Department of Veterans Affairs and of a Presidential Early Career Award for Scientists and Engineers (PECASE). This work was also supported by grants RO1 HD40789-01, RO1 HD38963-02, and R01-CA78052-01 from the National Institutes of Health, Bethesda, Md.

The authors have no relevant financial interest in this article.

We gratefully acknowledge Heidi Rininger for help with manuscript preparation and Allan Detsky, PhD, and Rodrigo Cavalcanti, PhD, for comments on an earlier draft.

REFERENCES

1. Vijan S, Hofer TP, Hayward RA. Cost-utility analysis of screening intervals for diabetic retinopathy in patients with type 2 diabetes mellitus. *JAMA*. 2000;283:889-896.
2. Newby KL, Eisenstein EL, Califf RM, et al. Cost effectiveness of early discharge after uncomplicated acute myocardial infarction. *N Engl J Med*. 2000;342:749-755.
3. Smith KJ, Roberts MS. The cost-effectiveness of sildenafil. *Ann Intern Med*. 2000;132:933-938.
4. Lerner M. Health insurers balk at paying $10 apiece for Viagra pills: Minnesota's three largest health plans say they need more time and details before they decide on whether—or to what extent—to pay. *Star Tribune*. June 3, 1998:1A.
5. Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a quality- adjusted life year: in search of a standard. *Med Decis Making*. 2000;20:332-342.
6. Neumann PJ, Sandberg EA, Bell CM, Stone PW, Chapman RH. Are pharmaceuticals cost-effective? a review of the evidence. *Health Aff (Millwood)*. 2000;19:92-109.
7. Ubel PA, Baron J, Nash B, Asch DA. Are preferences for equity over efficiency in health care allocation "all or nothing"? *Med Care*. 2000;38:366-373.
8. Ubel PA, Jepson C, Baron J, Hershey JC, Asch DA. The influence of cost-effectiveness information on physicians' cancer screening recommendations. *Soc Sci Med*. 2003;56:1727-1736.
9. Gold MR, Siegel J, Russell LB, Weinstein M, eds. *Cost-Effectiveness in Health and Medicine*. New York, NY: Oxford University Press; 1996.
10. Hershey J, Asch DA, Jepson C, Baron J, Ubel PA. Incremental and average cost-effectiveness ratios: will physicians make a distinction? *Risk Anal*. 2003;23:81-89.
11. Ritov II, Baron J. Protected values and omission bias. *Organ Behav Hum Decis Process*. 1999;79:79-94.
12. Eddy DM. Screening for cervical cancer. *Ann Intern Med*. 1990;113:214-226.
13. Fletcher S. Whither scientific deliberation in health policy recommendations? Alice in the Wonderland of breast-cancer screening. *N Engl J Med*. 1997;336:1180-1183.
14. Fendrick AM, Chernew ME, Hirth RA, Menon D. Understanding the behavioral response to medical innovation. *Am J Manag Care*. 1996;2:793-799.
15. Chernew ME, Fendrick AM, Hirth RA. Managed care and medical technology: implications for health care cost growth. *Health Aff (Millwood)*. 1997;16:196-206.
16. Eddy DM. Oregon's methods: did cost-effectiveness fail? *JAMA*. 1991;266:2135-2141.
17. Hadorn DC. Setting health care priorities in Oregon: cost-effectiveness meets the rule of rescue. *JAMA*. 1991;265:2218-2225.