Use of Cost-Effectiveness Analysis in Health-Care Resource Allocation Decision-Making: How Are Cost-Effectiveness Thresholds Expected to Emerge?

Hans-Georg Eichler, MD, MS,[1,2] Sheldon X. Kong, PhD,[2] William C. Gerth, MBA,[2] Panagiotis Mavros, PhD,[2] Bengt Jönsson, PhD[3]

[1]Vienna Center for Pharmaceutical Policy, Department of Clinical Pharmacology, Medical University of Vienna, Vienna, Austria; [2]Outcomes Research, Merck & Co., Whitehouse Station, NJ, USA; [3]Stockholm School of Economics, Stockholm, Sweden

ABSTRACT

Background: An increasing number of health-care systems, both public and private, such as managed-care organizations, are adopting results from cost-effectiveness (CE) analysis as one of the measures to inform decisions on allocation of health-care resources. It is expected that thresholds for CE ratios may be established for the acceptance of reimbursement or formulary listing.
Objective: This paper provides an overview of the development of and debate on CE thresholds, reviews threshold figures (i.e., cost per unit of health gain) currently proposed for or applied to resource-allocation decisions, and explores how thresholds may emerge.
Discussion: At the time of this review, there is no evidence from the literature that any health-care system has yet implemented explicit CE ratio thresholds. The fact

that some government agencies have utilized results from CE analysis in pricing/reimbursement decisions allows for retrospective analysis of the consistency of these decisions. As CE analysis becomes more widely utilized in assisting health-care decision-making, this may cause decision-makers to become increasingly consistent.
Conclusions: When CE analysis is conducted, well-established methodology should be used and transparency should be ensured. CE thresholds are expected to emerge in many countries, driven by the need for transparent and consistent decision-making. Future thresholds will likely be higher in most high-income countries than currently cited rules of thumb.
Keywords: cost-effectiveness, decision making, efficiency, health economics, health care, thresholds.

Introduction

Identifying the optimal allocation of available resources to maximize health will be the key challenge to health-care systems such as government agencies and managed-care organizations over the next decade. Medical research is expected to continue to produce an ever-increasing number of alternatives for the detection, prevention, and treatment of diseases. However, budgetary constraints will not allow health-care systems to make all of these available for everybody. This is probably recognized by health-care decision-makers in many countries, but their response to the challenge is, as yet, heterogeneous. Some have implemented an explicit or semi-explicit approach to guiding resource-allocation

decisions by formal health-economic analysis, the most popular approach currently being the cost-effectiveness (CE) analysis. This is frequently used in decision making in some countries, for example, Australia, Canada, Sweden, and the United Kingdom (UK). In most other countries, formal economic analysis is not yet a key input into the decision-making process [1]. However, there is an increasing awareness that resource allocation must be addressed in a systematic rather than intuitive manner. Several countries have recently introduced guidelines or legislation to mandate CE assessment of at least some aspects of health care, most often for the reimbursement of pharmaceuticals [2].

It is therefore reasonable to expect that decisions about resource allocation will increasingly rely on CE analysis. Inevitably, this will call for more transparency and consistency in the decision-making process and, in turn, for the definition of what policymakers regard as an "acceptable threshold" of cost-effectiveness below which they will make avail-

Address correspondence to: Hans-Georg Eichler, Department of Clinical Pharmacology, Medical University of Vienna, Allgemeines Krankenhaus, Währinger Gürtel 18-20, A-1090 Vienna, Austria. E-mail: hans-georg.eichler@meduniwien.ac.at

able a technology and above which they will ration access. It is recognized that a number of issues of cost-effectiveness assessment remain the subject of debate. However, this paper is not concerned with the technical limitations of CE analysis as such. We review the current concept of and debate on thresholds and discuss recent reports indicative of emerging CE thresholds. In the last section, we explore how thresholds are expected to evolve in future.

The Concept of Thresholds

The concept of "threshold" was originally proposed by Weinstein and Zeckhauser in 1973 [3], and refers to the level of costs and effects that an intervention must achieve to be acceptable in a given health-care system [4]. This implies a ratio between monetary cost, usually national currencies, the US dollar ($) or the European Euro (€), in the numerator and a measure of health gain in the denominator. The decision rule of CE analysis using a critical ratio results from a solution to a constrained optimization problem [3]. For mutually exclusive programs, CE ratios are usually and appropriately described as incremental CE ratios (ICER), comparing the cost-effectiveness of each intervention with that of the next most effective option [5].

Comparability of CE ratios is affected by the lack of a single, universally accepted measure of "health gain." A recent survey of health economic studies [6] demonstrated that cost-effectiveness analysis (CEA) and cost-utility analyses (CUA) are favored over cost–benefit analyses (CBA), but major differences remain in the choice of health-outcome measures across disease categories. However, Quality Adjusted Life-Years (QALYs), which represent years of healthy life weighted by a "utility" factor, have emerged as one of the most widely used and recommended measures by the academic community and many health-care systems [7,8]. In contrast, some international institutions, like the World Health Organization (WHO) or the World Bank frequently base their CE assessments on cost per Disability Adjusted Life-Year (DALY); [9,10], though the method of obtaining values for DALYs has been questioned [11]. Another measure of health gain that avoids the need for utility weighting is cost per Life-Year Gained (LYG) [12,13]. An in-depth discussion of some of the outcome measurements used in economic analyses can be found in Johannesson et al. [14]. Although the use of different denominators further complicates comparability of CE analysis-based decisions, most considerations about thresholds may be equally applicable to all types of

CE analyses, irrespective of the type of denominator used. In the following section, we briefly describe basic characteristics of CE thresholds.

Adoption of a threshold concept by a given health care system may be explicit or implicit. Here, explicit means that a group of decision-makers formally adopt and make public in advance any threshold (e.g., $/QALY) by which their decisions on resource allocation would be bound. In contrast, implicit thresholds are not official or public, but may be inferred retrospectively by analysis of the decision-making pattern in a given health-care system—provided there is at least some degree of consistency in decision making. The analysis by George et al. [13] of pharmaceutical reimbursement decisions in Australia offers an instructive example of implicit thresholds.

Setting an explicit threshold has been welcomed by a number of authors, as it offers a range of theoretical advantages [15]. These include reduced burden of responsibility upon those who previously made implicit rationing decisions alone, and better consistency and transparency of the decision-making process, equity, efficiency and public trust—as opposed to veiled denial of health-care services. Setting explicit thresholds would almost certainly generate public debate about societal willingness-to-pay for health care. Practical experience, such as that from the "Oregon experiment" [16] and results from public surveys [15,17,18] suggest that such debate might result in an increase in the health-care package and resource allocation to health care.

Setting explicit thresholds is politically sensitive, however, for these very reasons. In addition, decision-makers are not necessarily economists, and are reluctant to base their decisions on a single summary measure alone [19]. It is not entirely surprising therefore that, to the best of our knowledge, no single health-care system has yet implemented an explicit threshold. Not using explicit thresholds allows more room for arbitrariness and "ad hoc" considerations, which may be more attractive to policy decision-makers.

An additional characteristic of thresholds is their degree of flexibility, which distinguishes between hard and soft thresholds. A "hard" threshold approach dictates that results from CE analysis, expressed for example as $/QALY, are taken prima facie and become the sole decision criterion for resource allocation. While this rigid approach offers the theoretical advantages of transparency, consistency, and predictability, it denies the possibility of incorporating into the decision other, non-CE-based societal preferences. In contrast, adopting a "soft"

threshold would make room for considerations of other preferences. With this approach, the CE acceptability criterion does not lead to automatic acceptance or rejection but informs decision-making. Instead of a single figure, there is a threshold range with lower and upper boundaries (e.g., x–$1.5x$). R. Akehurst has metaphorically described this as "smudge" [20]. Interventions below the lower boundary will usually be accepted and made available, those above the upper level will usually be rejected, while the in-between "smudge [reflects a zone of] increasing discomfort as the cost goes higher and higher" [20]. This means that interventions falling between the lower and upper boundaries will be judged predominantly upon additional criteria.

Debates about the Threshold Concept

The idea of the use of thresholds in decision-making has met some criticism. Application of the critical threshold approach is only valid under a number of assumptions, including perfect divisibility of health-care programs, constant returns to scale, and constant marginal opportunity costs [21]. It was pointed out that these assumptions do not hold in the real world of health-care decision-making [21,22]. Moreover, it has been argued that the use of CE thresholds might lead to uncontrolled growth in health-care expenditure [23]. Application of CE thresholds alone ignores the fact that health-care systems are resource constrained and decision-makers need to balance their budgets. This is illustrated by a simple example given by Sendi et al. [24]: Program A with incremental costs of $100 producing 1 additional LYG has the same cost-effectiveness ratio as program B with incremental costs of $1000 and incremental effects of 10 LYG. Assume, both programs A and B fall below an accepted CE threshold, the number of patients eligible is substantial and is identical for both programs, and the programs, are essentially indivisible, e.g., due to high capital costs. It follows, that the total cost for program B is 10 times that for program A, and within a given budget, only program A but not program B may be affordable. This illustrates that in real-life situations, considerations of CE alone are insufficient to inform decision-makers. Decision-makers can either maximize health gain for a given budget, which gives an implicit CE ratio at the margin, or determine the budget based on an acceptable CE ratio. Both cannot be done at the same time [25]. However, the argument of constrained budgets does not invalidate the concept of CE thresholds [26].

Firstly, and in contrast to frequent rhetoric, health care budgets are not fixed, at least not in the long term. Reflecting on the Canadian health-care environment in 1992, Laupacis et al. [27] have argued "in a society as wealthy as ours, it is clear that if more health care funding was a societal priority, and if there was the political will, the available funds could still be increased." Ten years later this prediction was found to be true, not just in Canada. Health care budgets in the developed world have been rising over the past decade, both in absolute and relative (% of Gross Domestic Product [GDP]) terms (OECD Health Data 2002). A recent report on macroeconomics and health to the WHO [28] anticipates that even some of the lowest-income countries will increase their annual health outlays in the near-term future. This is not surprising, since both demographic changes and increasing societal reference points for what health care should achieve affect health outlays. To a large extent, this dynamic is driven by the availability of new treatments that change perceptions of what is "treatable disease" as opposed to "normal life" [29]. In the long run, the economic evaluation is expected to (and should) have an impact on the funds available for health care.

Secondly, a number of convincing theoretical arguments have been made in favor of using CE thresholds as opposed to a so-called health-care budget approach [26,30]. Most importantly from an ethical viewpoint, with a fixed-budget approach, decision-makers will tend to focus on their budget, while costs outside the budget would not be included. This approach has been termed "silo budgeting" and may occur within the health-care system and between health care and other areas of spending. Silo budgeting may be expected from decision-makers at the subsocietal level (e.g., hospital managers). However, it is not compatible with a societal perspective where analysts consider all health effects and costs that flow from an intervention, and where maximization of net social benefit per unit of constrained resource is the goal [31]. Hence, for societal decision-makers the CE-threshold decision rule seems more appropriate.

The relationship between CE thresholds and affordability in a constrained-budget situation is not straight forward. However, budget impact analysis can complement CE analysis [32,33]. Recently, Sendi and Briggs [24] have outlined how cost-effectiveness and affordability criteria for a given budget can be combined: this approach involves the construct of "cost-effectiveness affordability curves," based on an estimate of the number of

individuals that are candidates for treatment. Such a set of curves describes the probability that the treatment under consideration is both affordable and cost-effective as a function of the CE threshold and for different budget constraints.

Another controversy that has surrounded the application of CE analysis to the concept of thresholds, relates to the type of costs that should be included [34]. It has been argued that not just costs incurred due to the disease under consideration, but the total resource consequences resulting from any change in mortality must be taken into account [35]. Recently, Johannesson et al. [36] have shown how exclusion or inclusion of future costs affects CE ratios, particularly of interventions that increase life expectancy rather than improve quality of life. This illustrates the importance of prior definition of all costs to be included in CE analysis; where future costs are included, different thresholds may need to be applied.

League Tables and CE Thresholds
The threshold concept is not the only way of applying CE analysis to decision making. CE results for different interventions could be listed in descending order of cost-effectiveness, often called "league tables" [5]. In the presence of limited resources, programs can be implemented in sequence from the top down until a line is drawn where the budget is exhausted. Access is granted above, but not below the line, which moves up and down depending on the availability of resources [37]. The "league table" approach is attractive in theory because it combines the CE criterion with considerations of affordability, and all interventions, existing and new, are judged by the same standard, provided that issues of comparability can be overcome [5]. However, strict adherence to CE-based league tables, without consideration of issues of equity, may lead to anomalies [16] like allocation of an inordinate share of available resources to some diseases while leaving others untreated.

Moreover, the league table requires comprehensive information on the costs and effects of the complete menu of programs, which is not usually available. For that reason, a less data hungry alternative decision rule has been suggested: an existing program is identified that, if cancelled, would free up enough resources to fund the additional costs of a new program [38]. If the increased health outcomes associated with the new program are greater than the outcomes foregone from canceling the existing one, then the adoption of the new program represents a more efficient allocation of resources.

This approach was proposed as a second-best solution in that it can be used to identify improvements in, but not optimization of, resource allocations [38].

However, both the league-table and the "replacement" approach are poorly compatible with the managerial realities of health-care policies and politics. With few exceptions [16], health-care programs do not have the luxury of starting "anew," with a clean slate, but find they pay for an array of popular, traditional interventions, some of which may not be very cost-effective. Public willingness to forgo an existing program is generally lower than willingness to pay for a new program yielding the same benefit. This phenomenon has been fittingly described as a "kink in consumer's threshold value for cost-effectiveness in health care" [39]. As a consequence, and for psychological and political reasons, existing services are not simply at the wholesale disposal of decision-makers. Moreover, CE ratios are not available for many older interventions because they have not been evaluated. In reality, cost-effectiveness is most often assessed for emerging new technologies, on a one-by-one basis. In this situation, a combined threshold and budget impact approach appears more feasible than league tables.

Current Figures for Thresholds

In this section, we briefly review actual $/health gain threshold figures in the CE literature. We distinguish between threshold figures proposed by individual authors or institutions (rules of thumb), figures estimated from willingness to pay (WTP) or similar analysis, CE ratios obtained from other non-medical programs, and figures inferred from past health-care allocation decisions. The data discussed is summarized in Table 1.

Thresholds Proposed by Individuals or Institutions
In the US, a figure of US$50,000/QALY has frequently been quoted for many years as being cost-effective [40,41]. Hirth et al. [41] retrace how this number was originally based on the "dialysis standard," the purported annual cost/QALY to the Medicare program for patients with chronic renal failure. They estimate that, ironically, this widely cited standard might have been based on considerable underestimation of the program's true costs [41].

In 1992, Laupacis et al. [27] proposed that evidence for adoption of an intervention is strong if the CE ratio is 1990-CAN$20,000/QALY, moderate

Table I Summary of cost-effectiveness thresholds (cost/unit of health gain) and CE ratios that have been proposed for or applied to resource-allocation decisions. Thresholds are grouped by type of source

Reference	Country/description/methodology	Threshold as quoted in reference	Threshold converted/ inflated to 2002-US$
Thresholds proposed by individuals or institutions			
40,41	US, "rule of thumb," "dialysis standard"	US$50,000/QALY (quoted repeatedly since 1982)	93,500/QALY
27	Proposed in context of Canadian health-care system	l.b. 1990-CAN$20,000/QALY u.b. 1990-CAN$100,000/QALY	l.b. 17,600/QALY u.b. 87,800/QALY
42	US, summary of CE of cardiovascular interventions	l.b. 1996-US$40,000/LYG u.b. 1996-US$75,000/LYG	l.b. 44,800/LYG u.b. 83,900/LYG
43	Survey of UK and US health economists	1998(?)-US$60,000/LYG	65,000/LYG
45	Proposed for low-income countries	Less than 3 times GDP per capita per DALY averted	108,600/DALY (for US only; see also Table 2)
Thresholds estimated from willingness-to-pay or related studies			
41	NA + WE, human capital	1997-US$24,777/QALY	26,900/QALY
	NA + WE, revealed preference/non-occupational safety	1997-US$93,402/QALY	101,500/QALY
	NA + WE, contingent valuation	1997-US$161,305/QALY	175,300/QALY
	NA + WE, revealed preference/job risk	1997-US$428,286/QALY	645,400/QALY
50	UK, revealed preference	1997-UK£95,000/LYG	165,600/LYG
	UK, questionnaires	1997-UK£30,000/LYG	52,300/LYG
CE-ratios from other (nonmedical) programs			
12	US, health care	1993-US$19,000/LYG	23,000/LYG
	US, residential	1993-US$36,000/LYG	43,600/LYG
	US, transportation	1993-US$56,000/LYG	67,900/LYG
	US, occupational	1993-US$350,000/LYG	424,300/LYG
	US, environmental	1993-US$4,200,000/LYG	5,091,900/LYG
51	Sweden, medicine	1993-US$13,800/LYG	16,700/LYG
	Sweden, toxin control	1993-US$19,600/LYG	23,800/LYG
	Sweden, fatal injury reduction	1993-US$69,000/LYG	83,700/LYG
53	UK, road accident prevention	2002-UK£30,000/LYG	48,128/LYG
Thresholds inferred from past allocation decisions			
13	AUS, retrospective analysis of series of reimbursement decisions	l.b. 1998/9-AU$42,000/LYG u.b. 1998/9-AU$76,000/LYG	l.b. 28,200/LYG u.b. 51,000/LYG
2	NZ, assumption based on past reimbursement decisions	2000(?)-NZ$20,000/QALY	10,900/QALY
20	UK, retrospective analysis of past recommendations made by NICE	l.b. 2002(?)-UK£20,000/QALY u.b. 2002(?)-UK£30,000/QALY	l.b. 32,000/QALY u.b. 48,000/QALY

Abbreviations: CE, cost-effectiveness; DALY, disability-adjusted life-year; GDP, gross domestic product; l.b., lower boundary; LYG, life-year gained; NA + WE, North America and Western Europe; NICE, National Institute for Clinical Excellence (UK); QALY, quality-adjusted life-year; u.b., upper boundary.

if it is between CAN$20,000/QALY and CAN$100,000/QALY, and weak if it exceeds CAN$100,000/QALY. They acknowledged that these lower and upper boundaries were arbitrary but were in line with what had been "universally accepted as being appropriate" and what had been "provided routinely but . . . is significantly limited."

Goldman et al. [42] summarized their assessment of cost-effectiveness of a range of cardiovascular interventions with a recommendation: "At the current time[1996], programs that cost less than about $40,000 per year of life saved, which roughly corresponds to renal dialysis, have been recommended by some authors. Conversely, at costs above $75,0000 per year of life saved, we find it difficult to generate enthusiasm for an intervention . . ." No further justification is given in the article for the upper limit.

Newhouse [43] surveyed health economists about what threshold value to use in CE analysis, and reported a mean value of $60,000/LYS. Subsequent recommendations for treatment guidelines

were partly based on this finding [44]. Several more authors have suggested the adoption of different acceptable thresholds [6]. Most of these have two features in common: they are round numbers and they are rules of thumb in that there is no compelling rationale or justification for them. Discussing proposed threshold figures, Weinstein [40] pointed out with some irony that while they ignore inflation and currency exchange rates, "the appeal of the same round numbers is lasting."

The World Health Report 2002 [45] proposed a different approach to setting CE threshold. "The recent report of the Commission on Macroeconomics and Health, which was commissioned by WHO [28], suggested that interventions costing less than three times GDP per capita for each DALY averted represented good value for money." In the report of the Commission, this threshold is justified on the basis of expected direct and indirect benefits to national economies [28], though the report does not specify the types of costs that should be considered. This is remarkable for the intent to base allocation

Table 2 Theoretical values (in US$/DALY) for cost-effectiveness thresholds in several high-income countries, if thresholds were exclusively based on the "three times Gross Domestic Product (×3 GDP) per capita" approach proposed in the World Health Organization Report 2002 (WHO 2002). Values are based on Purchasing Power Parity-GDP per capita figures for 2000. (Source: The World Factbook 2001, accessed at http://www.bartleby.com/151/a64.html)

Country	"×3 GDP threshold" (US$/DALY)
USA	108,600
Japan	74,700
Canada	74,400
France	73,200
Germany	70,200
Australia	69,600
UK	68,400
Italy	66,300
Spain	54,000
New Zealand	53,100

DALY, Disability-Adjusted Life-Year.

decisions not on the appeal of arbitrary round numbers, but on an objective national benchmark that is directly related to the affordability criterion. Even though the recommendation was primarily directed towards low-income countries, it might also serve, with adjustments, as a yardstick for affluent health-care environments. Theoretical "×3 GDP thresholds" for select high-income countries are summarized in Table 2. Applying the unadjusted criterion, for instance, to the US (GDP per capita in 2000 around $36,200) this threshold would come to be in the order $108,600/DALY. It must be emphasized that the WHO recommendation cannot be directly compared to the $50,000/QALY (or similar) rules of thumb because the denominator is different. However, the concept of tying a national threshold to some type of objective economic benchmark, such as GDP, is readily adaptable to QALYs, or other denominators. Comparing QALY weights [46] to DALY weights [47] across a broad range of disease categories, it may be assumed that the "×3 GDP" WHO recommendation is well in excess of $50,000/QALY in many high-income countries.

Thresholds Estimated from Willingness-to-Pay or Related Studies

Suggestions from individuals or institutions or even analysis of past resource allocation decisions provide little guidance regarding the optimal CE-threshold that society is willing to accept. Therefore, obtaining more information about the willingness to pay (WTP) per unit of health gain in order to establish a useful decision rule has been a research priority for some time [48,49].

Theoretically, there are several avenues to arrive at a "societal $/QALY (or $/DALY)" value. These include the human capital approach, contingent valuation, also referred to as WTP, revealed preference/job risk, and revealed preference/nonoccupational safety. Hirth et al. [41] reviewed the literature for value-of-life estimates. They found wide variability in results, but showed that much of this was explained by methodology: median values (1997-US$) by study type were $24,777 (human capital), $93,402 (revealed preference/nonoccupational safety), $161,305 (contingent valuation), and $428,286 (revealed preference/job risk). The authors adjusted these estimates with age-specific quality-of-life weights and concluded that, with the exception of the human capital approach, most estimates were above the rule-of-thumb $50,000/QALY.

In their analysis, Hirth et al. also compared the implied value of a QALY across studies originating from different countries. They concluded that, at least for North America and Western Europe, cultural and economic differences do not seem to likely lead to different thresholds. However, while the analysis comprised 28 studies from the US, there were only 9 studies of non-US origin, from 4 different countries.

More recently, Hutton et al. [50] analyzed the value-of-life-year literature from the UK, and, combining values from a total of 12 studies and surveys, reported a median of 1997 UK£95,000/LYG from revealed preference estimates, and a median of £30,000/LYG from WTP questionnaires.

Attempts have been made to directly elicit the public's view on limitations of public health services for financial constraints. Rosen and Karlberg [18] report that 59% of a sample of Swedish citizens, but a much smaller fraction of politicians, administrators, and physicians fully agreed with the statement: "public health services should always offer the best possible care, irrespective of cost." Results from a similar UK-based survey are in broad agreement, 45% of the general public, but only 12% of clinicians disagreed that "there should be a limit on how much the National Health Service is allowed to spend" [17]. Such public views can hardly offer useful guidance for allocation of health-care budgets, but they do indicate that the public's valuation of any CE threshold may be higher than that of health-care policymakers. The National Institute for Clinical Excellence (NICE) in the UK is commissioning research to elicit the monetary value of health gains (e.g., expressed in QALYs) from the general population [8]. Results

from such studies will likely provide useful guidance for future CE thresholds.

CE Ratios from Other (Nonmedical) Programs

The cost-effectiveness of lifesaving interventions is an issue not only for health care but also for a number of other sectors such as transportation safety, occupational health and safety, environmental hazard control, fire prevention, etc. It is interesting to compare CE ratios from these sectors because the ratios from those programs that were actually funded could inform health-care policies about society's willingness to pay for health.

The most comprehensive compilation of CE ratios in the US for lifesaving interventions was undertaken by Tengs et al. [12], who listed the ICERs (expressed in 1993-US$/life-year saved) of 500 lifesaving interventions from different sectors. Not surprisingly, there was tremendous variation of ICERs over 11 orders of magnitude in almost every category (from net savings to more than $10 billions/LYS), and primary prevention was less cost-effective than secondary or tertiary prevention. More importantly, there were substantial differences in cost-effectiveness across sectors: Median ICER estimates by sector of society were $19,000/LYS (health care), $36,000/LYS (residential), $56,000/LYS (transportation), $350,000/LYS (occupational), and $4,200,000/LYS (environmental).

Ramsberg and Sjoberg [51] conducted a methodologically similar survey of 165 lifesaving interventions in Sweden. Their results were broadly comparable with those from the US [12], with the exception of environmental interventions. Median estimates of ICERs, expressed in 1993-US$/life-year saved, were: $13,800 (medical), $19,600 (toxin control), and $69,000 (fatal injury reduction). Viscusi and Aldy [52] compiled a comprehensive review of market estimates of the value of a statistical life from different countries and sectors; these results confirm the wide variability of estimates, but no attempt was made to convert these findings to ICERs per LYG or QALY.

Loomes [53] describes that in the UK, "at present, the values that are used by the highway authorities in appraising road traffic interventions are approximately £1 million for every expected fatality prevented . . . ," and, recalculating this figure, estimated it to represent approximately UK£30,000/LYG.

It must be emphasized that most of these estimates were based on $/life-years saved, which makes it difficult to compare them with $/QALY (or $/DALY) ratios. However, after inflating to 2002-$,

and if quality-of-life weights across a broad range of disease categories are imputed, most figures, including those for programs that have been implemented, are well in excess of the $50,000/QALY threshold for all sectors except health care.

Thresholds Inferred from Past Allocation Decisions

The most relevant information on implicit thresholds at work in a given health-care system can be gleaned from retrospective analysis of previous resource allocation decisions.

To our knowledge, the first systematic analysis of the track record of a health-care decision-making body was by George et al. [13], who retrospectively analyzed the decisions on reimbursement of drugs made by the Australian Pharmaceutical Benefits Advisory Committee (PBAC) between 1991 and 1996. The PBAC is charged with appraising, inter alia, the cost-effectiveness of new drugs and proposing a "recommend/reject" decision.

The authors were unable to identify a hard threshold beyond which the PBAC was unwilling to pay for additional health gain, indicating that the CE ratio was not the only factor determining the reimbursement decision. However, they found that the PBAC was unlikely to recommend a drug for reimbursement if the additional cost/LYG exceeded 1998/99-AU$76,000, approximately 1998/99-US$48,467, and was unlikely to reject a drug for which the additional cost/LYG was less than AU$42,000, approximately 1998/99-US$26,784. This result is consistent with the use of economic efficiency as a criterion for decision-making. If sustained in the future, it also illustrates the concept of a soft threshold with a reasonably well-defined lower and upper boundary, allowing for considerations of uncertainty, equity, or context of treatment. The lower and upper limits of the PBAC threshold corresponded to approximately 1.26 and 2.29 times Australian GDP per capita (US$21,200 in 1999). Note that these figures pertain to the "life-years gained" denominator of the CE ratio, which is unadjusted for quality of life. Due to the small number of PBAC decisions that were based on a $/QALY ratio, the authors could draw no conclusions about any QALY-based thresholds. However, if one were to impute a range of different quality weights [46], the lower and upper $/QALY boundaries would probably be in the order of just under 2 times and under 3 times GDP per capita.

Using a similar approach, Towse and Pritchard [20] analyzed the first 41 decisions made by the UK National Institute for Clinical Excellence (NICE) and conclude that, like the Australian PBAC,

this institution appears to operate a threshold range with a lower boundary of approximately UK£20,000/QALY, and an upper boundary of approximately UK£30,000/QALY. While NICE officials [54] have denied the existence of an "explicit threshold," this analysis by Towse and Pritchard appears convincing to the outside observer. If these figures should indeed evolve into the threshold range for the UK health-care system, the lower and upper boundaries would be in the order of 1.4 and 2.1 times GDP per capita, approximately US$22,800 in 2000. This is consistent with a low absolute and relative (% of GDP) level of health-care spending in the UK, as compared to other high-income countries (OECD Health Data 2002).

The Pharmaceutical Management Agency (PHARMAC) of New Zealand, established to make recommendations for the purchase of pharmaceuticals, has used CE analysis for a number of years. There is no published systematic analysis of its decision record, but at least one observer speculated that PHARMACs decisions are broadly consistent with a threshold of 2000-NZ$ 20,000/QALY [2].

CE analysis has been applied in several more countries, but there are only isolated reports of how other health-care systems arrived at individual reimbursement decisions based on cost per QALY [29,55]. At present, these examples are insufficient to infer implicit thresholds.

How Are Thresholds Expected to Evolve in Future?

Published opinions expressed by economists and policymakers on the concept of thresholds are split. However, we predict that CE thresholds will gradually become a reality, irrespective of whether local decision-makers welcome them or remain critical, because it is meaningless to perform CE studies in the absence of an acceptance threshold. Ironically, this evolution may follow the law of unintended consequences. As decision-makers progress to base resource allocations on CE analysis, it is possible to undertake retrospective systematic analysis of these decisions. Patient groups, providers of health care, and vendors of health-care technology have a predictably keen interest in tracking the consistency of the decision-making process—and to point out inconsistencies. It is difficult to see how this will not bind decision-makers to becoming increasingly consistent in their assessment of health-care technologies and eventually force a move from implicit to more explicit, predictable decision rules. It is anticipated, and should be welcomed, that this will not only lead to a convergence of decisions within, but also between defined sectors.

In spite of these considerations, neither theory nor empiric evidence supports the expectation that CE thresholds will evolve as the sole decision criterion in health-care resource allocation. While the adoption of a "hard" threshold theoretically guarantees a high level of efficiency of resource allocation, economic evaluations will have to be broadened to include other societal preferences, most importantly concerns of distributional equity. These concerns may become very relevant in some health-care environments. For example, Swedish politicians responsible for health care were prepared to sacrifice 15 out of 100 preventable deaths to achieve equity [56]. This has been described as the "equity–efficiency trade-off" [57], and relates also to the concept that a particular quantity of health gain (e.g., one QALY or DALY) accrued by different individuals (young vs. old, good vs. poor initial state of health) is not necessarily given the same weight [58]. Other concerns include preferential treatment given to specific disadvantaged groups of patients, e.g., those suffering from rare ("orphan") diseases or from acute, life-threatening diseases, often referred to as the "rule of rescue" [55]. Attempts have been made to operationalize decision rules that formally combine the CE criterion with other health-care priorities [59,60].

Based on these considerations we predict "soft" rather than "hard" thresholds to evolve, rather than being implemented by consensus or decree, as an economic decision support tool. Initially triggered by retrospective analysis of the decision track record, the development is expected to progress in most countries via an implicit to (eventually) an explicit stage. It may not be possible forever to "continue the myth that decisions about the allocation of care are based on clinical criteria alone" [61].

Conclusion

Explicit rationing is unpopular or actively discouraged [62], but implicit (and sometimes erratic) rationing balances budgets and maintains the system [63]. Over the past decade, an increasing number of health-care policymakers and managers have embraced health economics, and in particular CE analysis, as a tool for making allocation of resources more rational. The performance of such analyses and the appraisal of their results have direct cost, and the implicit costs of delaying access

to potentially beneficial health-care technologies [64]. Hence, conducting or requesting them makes sense only if their results are used in an appropriate way for allocation of scarce resources.

While CE analysis may have been intended to be a tool for cost-containment in the first place, this is not its proper use. Such studies may help detect underutilization as well as overutilization of health-care resources. What health-systems managers also may have overlooked, is the fact that CE analysis opens up to scrutiny their decision-making process, and the consistency of this process. The systematic analyses of the consistency of Australian and UK decision-making bodies [13,20] are but the first examples of things to come. In the face of public scrutiny, and a number of groups with strong economic interests in the allocation of healthcare resources, policymakers will be forced to become increasingly consistent. We cautiously predict that it is only a matter of time before this will move to a situation with explicit threshold values for cost-effectiveness. Considerations of cost-effectiveness will likely become only one of several criteria for resource allocation. Hence, the development of soft thresholds with upper and lower boundaries is more likely, and more sensible, than rigid implementation of a single CE criterion.

Over the past decades, the quality of clinical evidence has become the primary criterion for accepting and funding of health-care technology. Likewise, well-established methodology and valid and reliable data should be used and transparency should be ensured, when CE analysis is conducted. Moreover, economic efficiency in allocation of resources for health requires that the marginal health gain per $ spent be equal across investments. Hence, all interventions should logically be held up to the same standards of assessment and threshold levels. It has been pointed out that setting separate ceiling targets, e.g., for pharmaceutical expenditure, may not achieve the most efficient use of the overall health-care budget [2], and is difficult to defend.

Increased explicitness about health-care rationing is expected to lead to controversy, as rationing decisions are likely to be challenged at the individual-patient level and by interest groups or organizations [15,65]. The welcome consequence of this, for health-care providers and health-systems managers alike, is that health-care funding is also expected to increase as a result of public debate, when the political costs of open rationing become apparent. As a consequence, CE thresholds in most high-income countries may eventually be higher than currently cited rules of thumb. The general public's preparedness, at least in some countries, to increase resources to health care, even through higher taxation, is supported by preliminary evidence [17,66].

Such debate may also help redress allocation inefficiencies between lifesaving interventions in the health sector and other sectors [12]. We have discussed the substantial differences in cost-effectiveness of lifesaving interventions across sectors. Budget allocation decisions for the funded programs may or may not have been based on formal CE analysis, but may also have been driven by political priorities, environmental, or other concerns. Nonetheless, as Tengs et al. [12] pointed out, "this kind of variation is unnerving," because economic efficiency in promoting survival would also require that the marginal benefit per $ spent be equal across sectors.

It appears reasonable to expect that emerging thresholds will not be identical in different countries. The ability to pay for a given intervention varies with income level, even when costs and effectiveness are similar. This problem may also become apparent within the European Union (EU) as membership is extended to countries in Eastern Europe with much lower income levels. Different CE thresholds in the EU may give rise to tension and, perhaps, to the establishment of a health-equalization fund to defuse inter-country equity issues.

We hope that our discussion of factors influencing CE thresholds will help to point out more rational routes to arrive at a threshold than by perpetuating old rules of thumb. When discussing thresholds, it is time to say goodbye to the appeal of round numbers.

References

1 Ikegami N, Drummond M, Fukuhara S, et al. Why has the use of health economic evaluation in Japan lagged behind that in other developed countries? Pharmacoeconomics 2002;20(Suppl 2):S1–7.

2 Pritchard C. Overseas approaches to decision making. In: Towse A, Pritchard C, Devlin N, eds., Cost-Effectiveness Thresholds. Economic and Ethical Issues. London: King's Fund and Office of Health Economics, 2002.

3 Weinstein M, Zeckhauser R. Critical ratios and efficient allocation. J Public Econ 1973;2:147–57.

4 Torrance GW, Siegel JE, Luce BR. Framing and designing the cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds., Cost-Effectiveness in Health and Medicine.

Oxford, New York: Oxford University Press, 1996.

5 Gold MR, Siegel JE, Russell LB, Weinstein MC, eds., Cost-Effectiveness in Health and Medicine. Oxford, New York: Oxford University Press, 1996.

6 Anell A, Norinder A. Health outcome measures used in cost-effectiveness studies: a review of original articles published between 1986 and 1996. Health Policy 2000;51:87–99.

7 Weinstein MC, Siegel JE, Gold MR, et al. Recommendations of the panel on cost-effectiveness in health and medicine. JAMA 1996;276:1253–8.

8 Advertisement: NHS R & D Research Methodology Programme and the National Institute for Clinical Excellence (NICE): RM03/JH12: The societal value of health gains. Available from: http://www.publichealth.bham.ac.uk/nccrm/invitations_to_tender.htm [Last accessed February 8, 2004].

9 Raisch DW. Understanding quality-adjusted life years and their application to pharmacoeconomic research. Ann Pharmacother 2000;34:906–14.

10 Melse JM, Essink-Bot ML, Kramers PGN, Hoeymans N. A national burden of disease calculation: Dutch disability-adjusted life years. Am J Public Health 2000;90:1241–7.

11 Williams A. Calculating the global burden of disease: time for a strategic reappraisal. Health Econ 1999;8:1–8.

12 Tengs TO, Adams ME, Pliskin JS, et al. Five-hundred life-saving interventions and their cost-effectiveness. Risk Anal 1995;15:369–90.

13 George B, Harris A, Mitchell A. Cost-effectiveness analysis and the consistency of decision making. Evidence from pharmaceutical reimbursement in Australia (1991–1996). Pharmacoeconomics 2001;19:1103–9.

14 Johannesson M, Jonsson B, Karlsson G. Outcome measurement in economic evaluation. Health Econ 1996;5:279–96.

15 Coast J. Who wants to know if their care is rationed? Views of citizens and service informants. Health Expect 2001;4:243–52.

16 Ham C. Retracing the Oregon trail: the experience of rationing and the Oregon health plan. BMJ 1998;316:1965–9.

17 Lees A, Scott N, Scott SN, et al. Deciding how NHS money is spent: a survey of general public and medical views. Health Expect 2002;5:47–54.

18 Rosen P, Karlberg I. Opinions of Swedish citizens, health-care politicians, administrators and doctors on rationing and health-care financing. Health Expect 2002;5:148–55.

19 Zwart-van Rijkom JEF, Leufkens HGM, Busschbach JJV, et al. Differences in attitudes, knowledge and use of economic evaluations in decision-making in The Netherlands. The Dutch results from the EUROMET project. Pharmacoeconomics 2000; 18:149–60.

20 Towse A, Pritchard C. Does NICE have a threshold? An external view. In: Towse A, Pritchard C, Devlin N, eds., Cost-Effectiveness Thresholds. Economic and Ethical Issues. London: King's Fund and Office of Health Economics, 2002.

21 Birch S, Gafni A. Changing the problem to fit the solution: Johannesson and Weinstein's (mis) application of economics to real world problems. J Health Econ 1993;12:469–76.

22 Sendi P, Al MJ. Revisiting the decision rule of cost-effectiveness analysis under certainty and uncertainty. Soc Sci Med 2003;57:969–74.

23 Gafni A, Birch S. Guidelines for the adoption of new technologies: a prescription for uncontrolled growth in expenditures and how to avoid the problem. Can Med Assoc J 1993;148:913–7.

24 Sendi PP, Briggs AH. Affordability and cost-effectiveness: decision-making on the cost-effectiveness plane. Health Econ 2001;10:675–80.

25 Karlsson G, Johannesson M. The decision rules of cost-effectiveness analysis. Pharmacoeconomics 1996;9:113–20.

26 Garber AM, Phleps CE. Economic foundations of cost-effectiveness analysis. J Health Econ 1997; 16:1–31.

27 Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. Can Med Assoc J 1992;146:473–81.

28 WHO Commission on Macroeconomics and Health. Macroeconomics and health: investing in health for economic development. Report of the Commission on Macroeconomics and Health. Geneva: World Health Organization, 2001.

29 Stolk EA, Brouwer WBF, Busschbach JJV. Rationalising rationing: economic and other considerations in the debate about funding of Viagra. Health Policy 2002;59:53–63.

30 Johannesson M, O'Connor RM. Cost-utility analysis from a societal perspective. Health Policy 1997;39:241–53.

31 Russell LB, Gold MR, Siegel JE, et al. The role of cost-effectiveness analysis in health and medicine. JAMA 1996;276:1172–7.

32 Trueman P, Drummond M, Hutton J. Development guideline for budget impact analysis. Pharmacoeconomics 2001;19:609–21.

33 Nuijten MJ, Rutten F. Combining a budgetary-impact analysis and a cost-effectiveness analysis using decision-analytic modelling techniques. Pharmacoeconomics 2002;20:855–67.

34 Olsen JA, Richardson J. Production costs from health care: what should be included in cost-effectiveness analyses? Soc Sci Med 1999;49:17–26.

35 Johannesson M. The concept of cost in the economic evaluation of health care. Int J Technol Assess Health Care 1994;10:675–82.

36 Johannesson M, Meltzer D, O'Connor RM. Incorporating future costs in medical cost-effectiveness analysis: implications for the cost-effectiveness of the treatment of hypertension. Med Decis Making 1997;17:382–9.

37 New B. Defining a package of healthcare services the NHS is responsible for. The case for. BMJ 1997;314:503–5.

38 Sendi P, Gafni A, Birch S. Opportunity costs and uncertainty in the economic evaluation of health care interventions. Health Econ 2002;11: 23–31.

39 O'Brien BJ, Gertsen K, Willan AR, Faulkner LA. Is there a kink in consumers' threshold value for cost-effectiveness in health care? Health Econ 2002;11:175–80.

40 Weinstein MC. From cost-effectiveness ratios to resource allocation: where to draw the line? In: Sloan FA, ed., Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies. New York: Cambridge University Press, 1995.

41 Hirth RA, Chernew ME, Miller E, et al. Willingness to pay for a quality-adjusted life year: in search of a standard. Med Decis Making 2000;20:332–42.

42 Goldman L, Garber AM, Grover SA, Hlatky MA. Task force 6. Cost effectiveness of assessment and management of risk factors. J Am Coll Cardiol 1996;27:964–1047.

43 Newhouse JP. US and UK health economics: two disciplines separated by a common language? Health Econ 1998;7(Suppl):S79–92.

44 Johannesson M. At what coronary risk level is it cost-effective to initiate cholesterol lowering drug treatment in primary prevention? Eur Heart J 2001;22:919–25.

45 World Health Organization. World Health Report 2002. Geneva: World Health Organization, 2002.

46 Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. Med Care 2000;38:583–637.

47 Murray CJL, Lopez AD. The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries and Risk Factors in 1990 and Projected to 2020. Cambridge, MA: Harvard University Press, 1996.

48 Card WI, Mooney GH. What is the monetary value of a human life? BMJ 1977;2:1627–9.

49 Johannesson M, Meltzer D. Some reflections on cost-effectiveness analysis. Health Econ 1998;7:1–7.

50 Hutton J, Benedict A. (MEDTAP® International, London UK) Producing Health Benefits: Cost-Effectiveness of Health and Non-Health Interventions. Presented at Clinical Excellence: Annual Conference and Exhibition, December 3–5, 2002, Birmingham, UK.

51 Ramsberg JAK, Sjoberg L. The cost-effectiveness of lifesaving interventions in Sweden. Risk Anal 1997;17:467–78.

52 Viscusi WK, Aldy JE. The value of a statistical life: a critical review of market estimates throughout the world. Working paper 9487, National Bureau of Economic Research, February 2003. Last accessed at http://www.nber.org/papers/w9487.

53 Loomes G. Valuing life years and QALYs: transferability and convertibility of values across the UK public sector. In: Towse A, Pritchard C, Devlin N, eds., Cost-Effectiveness Thresholds. Economic and Ethical Issues. London: King's Fund and Office of Health Economics, 2002.

54 Littlejohns P. Does NICE have a threshold? A response. In: Towse A, Pritchard C, Devlin N, eds., Cost-Effectiveness Thresholds. Economic and Ethical Issues. London: King's Fund and Office of Health Economics, 2002.

55 Gross ML. Ethics, policy, and rare genetic disorders: the case of Gaucher disease in Israel. Theor Med Bioeth 2002;23:151–70.

56 Lindholm L, Rosen M, Emmelin M. How many lives is equity worth? A proposal for equity adjusted years of life saved. J Epidemiol Community Health 1998;52:808–11.

57 Wagstaff A. QALYs and the equity-efficiency trade-off. J Health Econ 1991;10:21–41.

58 Schwappach DLB. Resource allocation, social values and the QALY: a review of the debate and empirical evidence. Health Expect 2002;5:210–22.

59 Government Committee on Choices in Health Care. Choices in Health Care. Zoetermeer: Ministry of Welfare, Health and Cultural Affairs, 1992.

60 Stolk E, Bussschbach J. Economics and ethics in health care. Where can they meet? In: Gastmans C, ed., Between Technology and Humanity. The Impact of Technology on Health Care Ethics. Leuven, Belgium: Leuven University Press, 2002.

61 Doyal L. Rationing within the NHS should be explicit. The case for. BMJ 1997;314:1114–18.

62 Klein R. Defining a package of healthcare services the NHS is responsible for. The case against. BMJ 1997;314:506–9.

63 Mcdonald R. Street-level bureaucrats? Heart disease, health economics and policy in a primary care group. Health Soc Care Community 2002;10:129–35.

64 Edmonds P, Glynn D, Oglialoro C. Access to important new medicines. European Business J 2000;12:146–57.

65 Anonymous. MS Treatment and NICE. Parliamentary Office of Science and Technology. Postnote February 2002 Number 168. Available from: http://www.parliament.uk/post/home.htm.

66 Bosanquet N, Zarzecka A. Attitudes to health services 1983 to 1993. In: Harrison A, ed., Health Care UK 1994/95. London: King's Fund Policy Institute, 1995.