Thresholds for the cost–effectiveness of interventions: alternative approaches

Elliot Marselle, Bruce Larson, Dhruv S Kazi, James G Kahn & Sydney Rosen

Abstract Many countries use the cost–effectiveness thresholds recommended by the World Health Organization’s Choosing Interventions that are Cost–Effective project (WHO-CHOICE) when evaluating health interventions. This project sets the threshold for cost–effectiveness as the cost of the intervention per disability-adjusted life-year (DALY) averted less than three times the country’s annual gross domestic product (GDP) per capita. Highly cost–effective interventions are defined as meeting a threshold per DALY averted of once the annual GDP per capita. We argue that reliance on these thresholds reduces the value of cost–effectiveness analyses and makes such analyses too blunt to be useful for most decision-making in the field of public health. Use of these thresholds has little theoretical justification, skirts the difficult but necessary ranking of the relative values of locally-applicable interventions and omits any consideration of what is truly affordable. The WHO-CHOICE thresholds set such a low bar for cost–effectiveness that very few interventions with evidence of efficacy can be ruled out. The thresholds have little value in assessing the trade-offs that decision-makers must confront. We present alternative approaches for applying cost–effectiveness criteria to choices in the allocation of health-care resources.

Introduction

In public health, cost–effectiveness analyses compare the costs and effectiveness of two or more health interventions – with effectiveness measured in the same units. When comparing interventions, the incremental cost–effectiveness ratio (ICER) – i.e. the difference in costs divided by the difference in health effects – is often used to express the result.

Estimates of costs, health effects and ICERs provide clear guidance to policy-makers in three situations: (i) when the health-effect target is specified by policy-makers and the aim of the cost–effectiveness analysis is to minimize the expenditure needed to achieve that target; (ii) when a budget constraint is specified by policy-makers and the aim is to maximize the health benefits while keeping expenditure within budget; and (iii) when policy-makers have specified an explicit standard or threshold for what should be considered cost–effective.

In all three cases, the analysts completing the cost–effectiveness analysis cannot objectively make a recommendation to policy-makers without prior decisions by policy-makers on health-effect or cost targets or thresholds. Without reference to such decisions, the cost–effectiveness analysis cannot fully orient policy-makers to the range of options that might be good investments.

For example, compared with no vaccination, routine quadrivalent human papillomavirus vaccination combined with catch-up vaccination – to protect against cervical diseases in Brazil – was found to have an ICER of 450 United States dollars (US$) per quality-adjusted life-year (QALY) gained. In the United Republic of Tanzania, compared with no treatment, post-exposure prophylaxis for rabies was found to have an estimated ICER of US$ 27 per QALY gained. However, how does one decide whether US$ 450 per QALY gained in Brazil or US$ 27 per QALY gained in the United Republic of Tanzania represents good use of money for the national health-care system?

Three general approaches have been used to solve this problem: (i) thresholds based on per capita national incomes; (ii) benchmark interventions and (iii) league tables. In recent years, the most common approach has involved the use of thresholds based on per capita gross domestic product (GDP). Under this approach – which has been promoted by the World Health Organization’s Choosing Interventions that are Cost–Effective (WHO-CHOICE) project – an intervention that, per disability-adjusted life-year (DALY) avoided, costs less than three times the national annual GDP per capita is considered cost–effective, whereas one that costs less than once the national annual GDP per capita is considered highly cost–effective.

In this article, we argue that the current thresholds based on per capita GDP have major shortcomings as guides for policy-makers, since each of the available approaches has substantial weaknesses. We then discuss that a new consensus should be reached on a process for evaluating the cost–effectiveness of health interventions that places ICERs in the context of other, local policy and programme options, including funding sources. We focus on cost–effectiveness and ignore other criteria for policy decisions, such as equity, ethics and political feasibility. We proceed from the premise that evidence-based economic evaluations are vital additions to public policy decisions – which would otherwise largely reflect political, ideological and/or bureaucratic concerns. We focus on the relative merits of different ways of distinguishing what constitutes an acceptable level of cost–effectiveness and on the limitations of the widely used national-income-based approach.

References

1 Health Strategies International, 555 Fifty-ninth Street, Oakland, California, 94609, United States of America (USA).
2 Center for Global Health and Development, Boston University, Boston, USA.
3 Division of Cardiology, San Francisco General Hospital, San Francisco, USA.
4 Institute for Health Policy Studies, University of California – San Francisco, San Francisco, USA.
5 Correspondence to Elliot Marselle (email: emarselle@comcast.net).
6 Submitted: 5 March 2014 – Revised version received: 27 October 2014 – Accepted: 26 November 2014 – Published online: 15 December 2014.

doi: http://dx.doi.org/10.2471/BLT.14.138206
Thresholds

The most pervasive threshold was initially promoted by the Commission on Macroeconomics and Health and adopted in The world health report 2002 and by WHO-CHOICE. This threshold links per capita GDP with returns on investments in health to define the characteristics of a cost–effective and a very cost–effective intervention.4–6 Many published cost–effectiveness analyses of health interventions in low resource countries now explicitly refer to these WHO criteria as the standards by which each intervention is considered cost–effective or not. However, use of these criteria has at least four major limitations.

The first limitation is that important comparisons are obscured. Cost–effectiveness analysis is useful only in the context of the choices available in a particular setting and context – e.g. the budget and technical capacity of a national malaria control programme or Ministry of Health. Even if an intervention is categorized as cost–effective based on its cost per DALY averted, that intervention may still not represent the best use of a country’s health budget (Box 1). It is not enough to know that, per DALY avoided, an intervention costs less than three times the local annual per capita gross domestic product. We also need to know if it costs less – per DALY avoided – than other needed and feasible interventions. The current shift in some of the United States of America’s global health funding – i.e. away from support for the treatment of human immunodeficiency virus (HIV) infections and towards malaria, maternal and child health and other programmes – tacitly recognizes that, even among activities with ICERs below a national-income threshold, trade-offs are real and consequential.

The second limitation is that thresholds are too easily attained. Beyond the virtue of availability, we are puzzled why per capita gross domestic products were chosen as the main units for cost–effectiveness thresholds. Too many health interventions are found to cost less, per DALY averted, than the relevant annual per capita gross domestic product. Box 2 illustrates this problem for diarrhoeal disease control. Making the threshold harder to meet – e.g. by only categorizing an intervention as highly cost–effective if, per DALY averted, it costs less than half of the annual per capita GDP – does not address the fundamental problem, which is that any threshold is arbitrary. More stringent thresholds would rule interventions out with as little justification as more lenient thresholds would rule them in.

The third limitation is the untested assumptions on which this approach is based. Social willingness to pay for health benefits is, conceptually, an appropriate way to define social value that could be informed by the results of non-market valuations based on revealed- and stated-preference approaches.12 In using a cost–effectiveness threshold that is based on a country’s per capita GDP, analysts tacitly assume that the country is willing to pay up to that threshold for the health benefit – usually without any concrete evidence of that willingness to pay. While willingness to pay for health care is related to income, there is little evidence that the relationship is linear. Other factors are also important. If averted DALYs are more highly valued in high-income countries than in low-income ones,20 use of cost–effectiveness thresholds based on per capita GDP per DALY averted will give a biased measure of the willingness to pay. Such thresholds will tend to be too stringent in high-income countries – thus ruling some efficient options in. Some inefficient options out – and too lax in low-income countries – thus ruling some inefficient options in.

The fourth limitation is that affordability is not adequately appraised. Cost–effectiveness analyses are typically addressed to governments or international donors and aim to assist decision-making about how to spend finite budgets. Recent experience with international funding for HIV programmes may have fostered
the notion that budget constraints are illusory. However, even HIV funding is less secure now than it was a few years ago.\textsuperscript{21–23} There is no evidence that, in the short term at least, the world will contribute the sums needed to implement all interventions that meet WHO’s criteria for cost–effectiveness. Thus, in any timeframe relevant to policy-makers, trade-offs have to be considered.

Ignoring the overall budget assigned to a health programme may be just as problematic in a high-income country as in a lower-income one – particularly for conditions that are highly prevalent. Consider a drug that adds a year to everyone’s life and costs the annual per capita GDP per person treated. Although such a drug would be categorized as highly cost–effective by WHO’s thresholds, we would have to spend the entire GDP of the country each year to give the drug to every eligible individual – i.e. to the country’s entire population.

### Benchmark interventions

Originally proposed by Weinstein and Zeckhauser,\textsuperscript{26} a second solution to the cost–effectiveness standard problem is to cite the cost–effectiveness of a benchmark intervention that has already been adopted in the relevant country and to use that as a threshold for acceptable cost–effectiveness. In this approach we are again using a threshold but – unlike the thresholds based on per capita GDP – this threshold is established by a retrospective analysis of existing practice.\textsuperscript{27} In the USA, for example, a threshold still used in cost–effectiveness analyses – US$ 50,000 per QALY gained – was based on an estimate of the cost–effectiveness of dialysis for chronic renal disease.\textsuperscript{19} This threshold has recently been updated to US$ 100,000 or US$ 150,000 per QALY gained.\textsuperscript{28} Since there is already evidence of a willingness to pay US$ 150,000 per QALY gained, it should be possible to increase overall health benefits by transferring funds from activities that cost more than this sum to activities that cost less. Thus, this approach appears to justify the adoption of any option that has a lower ICER than the benchmark.

Although such an approach may have better local relevance than thresholds based on per capita GDP, it also has substantial shortcomings. The ICER of the benchmark intervention may be a high or low outlier. For example, it may have resulted from a political decision that does not reflect the current, true measure of societal willingness to pay for health benefits. In addition, benchmarks do not take affordability into account and are not regularly updated to reflect changes in opportunity costs resulting from new technologies or delivery models, or changes in the burden of disease.

Most importantly, using a single benchmark does not address the critical question of whether there might be available options that have a better cost–effectiveness ratio than either the benchmark intervention or the intervention under evaluation. In the USA, for example, an analysis might reveal that an intervention can add a QALY for US$ 80 000 – i.e. well under the US$ 150 000 benchmark cited above. Although this would indicate that the intervention is much more cost–effective than the current benchmark, it would not tell us anything about the set of possible interventions that might add a QALY for less than US$ 80 000. Other techniques for establishing thresholds, such as human capital, contingent valuation and revealed preference approaches\textsuperscript{29} share the same basic strengths and weaknesses as the benchmark approach. An option to justify the one under study can almost always be found.\textsuperscript{30,31} One way to mitigate this problem is to consider a range of interventions adopted by public health programmes in the setting of interest and the range of ICERs from these adopted interventions. This could be achieved via a research agenda that aims to aggregate more data on willingness to pay for a unit of health benefit in a wide range of countries. In high-income countries, progress has been made on such an agenda by the translation of the available data on lives saved to data on QALYs gained.\textsuperscript{32}

### League tables

A third approach side-steps the threshold question and focuses instead on getting the largest health impact for the budget. Conceptually, a complete set of relevant interventions would be chosen to maximize health effects. For example, if all of the interventions considered are at least somewhat scalable, they can be ranked into a so-called league table according to their ICERs.\textsuperscript{33} The league-table approach is based on the principle that, for any budget, health outcomes are maximized if selection of the options for implementation begins at the top of the league table – i.e. with the option with the lowest ICER – and then moves down the list, to interventions with successively higher ratios, until the budget is exhausted.\textsuperscript{31}

Several generic league tables have been developed. WHO-CHOICE has reported simple information on the ICERs for many interventions.\textsuperscript{3} Separate regional league tables are available for several diseases or risk factors. For example, for the Africa D region there are tables for 60 different interventions (Table 1). Other league tables have been created for specific diseases or conditions. A 2005 article assessed the ICERs of several major HIV interventions and arranged these in a league table for sub-Saharan Africa and South-East Asia (Table 2).\textsuperscript{32} Other league tables are large repositories of cost–effectiveness information that can be used to assess the ranking of many interventions for wide ranges of diseases and conditions. One of the largest of these is the cost–effectiveness analysis registry maintained by Tufts Medical Center, which provides over 3600 ICERs for over 2000 health interventions.\textsuperscript{34}

A limitation of league tables is that ICERs may not be available for many relevant options or settings. Many low resource countries lack data on the costs and effectiveness of specific interventions. In these countries, the only recourse for local policy-makers is to use findings from similar countries. A bare league table omits much of the information that decision-makers might want to consider when choosing among options – e.g. the size of the affected population, whether the intervention is scalable, the health benefit per recipient and the degree of uncertainty around the ICERs.\textsuperscript{35,36} Perhaps, given these, we need an extended league table approach in which a list of ICERs is complemented by information on context-sensitive costs and benefits of competing options.

Against these disadvantages must be weighed several virtues. A league table indicates graduated distinctions between ICERs. Since the length of the list of interventions deemed cost–effective varies according to the budget, league tables combine considerations of cost–effectiveness with affordability.\textsuperscript{37} The last (least cost–effective) intervention in the table to be adopted is more likely to approximate society’s willing-
ness to pay for health benefits than the open-ended set of commitments implied by global thresholds. Finally, league tables need not be comprehensive to support improved resource allocation. They can still indicate the potential health benefits of cancelling an existing programme and using the resources freed to fund another programme.
Cost–effectiveness thresholds

Elliott Marseille et al.

Policy & practice

Abstract

Many countries in assessing health interventions use World Health Organization thresholds for cost–effectiveness (CHOICE).的选择具有成本效益的干预措施项目

They use many of these thresholds a single year’s potential GDP avoids the same years of healthy life lost from disability adjusted life years (DALYs). We argue that the estimates of both costs and effectiveness must be situated firmly in the relevant context, which includes the disease burden and budget of the setting in question. Simple cost–effectiveness thresholds – whether based on per-capita incomes or benchmark interventions – fail to evaluate and rank interventions within countries and disregard budgetary constraints. By short-circuiting a more thorough assessment of policy-relevant alternatives, they contribute little to good decision-making and can actually mislead. While the currently available data will not support comprehensive off-the-shelf league tables for most settings, the results of cost–effectiveness analyses should be compared with as many relevant interventions as reasonable in a given situation. Decision-makers would then be in a far better position to interpret the results of cost–effectiveness analyses.

A consensus process should be convened, perhaps by WHO, to develop a new framework for articulating cost–effectiveness in global health policy – specifically focusing on low- and middle-income countries. Rather than referencing a uniform standard, this new consensus should place ICERs in the context of other public health options available or already adopted in the relevant country setting – and in the context of the relevant budgets. While not resolving all of the issues affecting cost–effectiveness analysis as a guide for resource allocation, a new framework could offer an improvement on the use of simple thresholds based on per-capita incomes.

Competing interests: None declared.
