CHE
Centre For Health Economics

UNIVERSITY of York

Cost-Effectiveness Thresholds in Health Care: A Bookshelf Guide to their Meaning and Use

Anthony J Culyer

CHE Research Paper 121

Cost-effectiveness thresholds in health care: a bookshelf guide to their meaning and use

Anthony J Culyer

Emeritus Professor, Centre for Health Economics, University of York, UK
Adjunct Professor, University of Toronto, Canada

December 2015

Background to series

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Acknowledgements

An earlier version of this paper was given as the 2015 Emmet Hall Lecture at the Montreal meeting of the Canadian Association of Health Services and Policy Research (CAHSPR). My thanks for comments on this and preceding papers with a closely related theme are due to: Kalipso Chalkidou, Karl Claxton, Richard Edlin, Karen Hofman, Chris McCabe, James F O'Mahony, Adam Oliver, Mark Sculpher, Mike Paulden, Paul Revill, Eldon Spackman, Yot Teerawattananon, Adrian Towse and Walter Wodchis.

Centre for Health Economics
Alcuin College
University of York
York, UK
www.york.ac.uk/che

Abstract

There is misunderstanding about both the meaning and the role of cost-effectiveness thresholds in policy decision making. This article dissects the main issues by use of a bookshelf metaphor. Its main conclusions are these:

- It must be possible to compare interventions in terms of their impact on a common measure of health.
- Mere effectiveness is not a persuasive case for inclusion in public insurance plans.
- Public health advocates need to address issues of relative effectiveness.
- A 'first best' benchmark or threshold ratio of health gain to expenditure identifies the least effective intervention that should be included in a public insurance plan.
- The reciprocal of this ratio – the 'first best' cost-effectiveness threshold – will rise or fall as the health budget rises or falls (ceteris paribus).
- Setting thresholds too high or too low costs lives.
- Failure to set any cost-effectiveness threshold at all also involves avertable deaths and morbidity.
- The threshold cannot be set independently of the health budget.
- The threshold can be approached from either the demand-side or the supply side – the two are equivalent only in a health-maximising equilibrium.
- The supply-side approach generates an estimate of a 'second best' cost-effectiveness threshold that is higher than the 'first best'.
- The second best threshold is the one generally to be preferred in decisions about adding or subtracting interventions in an established public insurance package.
- Multiple thresholds are implied by systems having distinct and separable health budgets.
- Disinvestment involves eliminating *effective* technologies from the insured bundle.
- Differential (positive) weighting of beneficiaries' health gains may increase the threshold.
- Anonymity and identity are factors that may affect the interpretation of the threshold.
- The true opportunity cost of health care in a community, where the effectiveness of interventions is determined by their impact on health, is not to be measured in money – but in health itself.

Introduction

The idea of having a cost-effectiveness ratio as a guide for selecting health care interventions for inclusion in a national health insurance scheme has proved controversial. It has a long history (e.g. Weinstein and Zeckhauser 1973, Weinstein and Stason 1977, Doubilet 1986, Birch and Gafni 1992, Laupacis et al. 1992, George et al. 2001, Towse et al. 2002, Devlin and Parkin 2004, Gafni and Birch 2006, Culyer et al. 2007, Chambers et al. 2010, Eckermann and Pekarsky 2014, Newall et al. 2014, Danzon et al. 2015a and b, Marseille et al. 2015, Culyer 2015). It is in practical use in some countries (e.g. Australia, England and Wales). It is advocated by significant agencies in wealthy as well as low- and middle-income countries (LMICs) (e.g. Office of Health Economics, World Health Organisation) and is dismissed as arbitrary and politically repressive by others (Nyhan 2010). There is misunderstanding about both its meaning and its role in decision making.

This article explains the essential meaning of a cost-effectiveness threshold, using the simple metaphor of a bookshelf. Implications of its use, misuse and non-use are explored. These matters are discussed in the context of decisions by governments and agencies concerning the inclusion or exclusion of healthcare technologies in public programmes. To keep matters simple, the assumptions made throughout are that insured persons have access to technologies free of charge and that the principal objective of such schemes is to promote population health. The assumption is maintained throughout that "effectiveness" relates to the impact of an intervention on people's health. This is not to deny that health care systems may have objectives other than health maximisation but to focus on what is undoubtedly a major objective[1]. The analysis is considered in the context of countries of varying degrees of economic development but the main focus is on LMICs considering how best to advance universal health coverage (UHC) by introducing a public health insurance scheme (PHI).

What treatments should be included in public insurance schemes?

The World Health Organisation (WHO) defines as 'essential' "those drugs that satisfy the health care needs of the majority of the population; they should therefore be available at all times in adequate amounts and in appropriate dosage forms, at a price the community can afford" (WHO 2003). This is a curious definition, partly because it contains a value judgment to the effect that such drugs ought to be provided and an ambiguous condition that they be provided only at a price the community in question can afford. In practice the WHO has a long list of 'essential' drugs but leaves it up to local 'communities' to determine which ones are to feature on their local list. In practice, then, whether a medicine is 'essential', and therefore provided at all times in adequate amounts and appropriate dosages, depends on whether a local community chooses to afford it.

One such community is Tanzania, whose current essential list contains more than 500 medicines with many controversial drugs on it such as Avastin in addition to Taxol and Paraplatin for treating ovarian cancer, and Lucentis for treating macular eye disease (Tanzania Ministry of Health and Social Welfare 2013, p 279). The first of these is regarded as not cost-effective by NICE in England and Wales at approximately £144,000 ($206,000) per Quality-Adjusted Life-Year (QALY) and the second is recommended only if the manufacturer offers substantial discounts (NICE 2013 a, b). They stand, nonetheless, on Tanzania's list of cost-effective 'essential' medicines.

[1] Other objectives commonly include financial protection (e.g. from the out-of-pocket expense of costly interventions), reassurance, information provision (e.g. diagnostic utility), certification (e.g. for legitimate absence from work), reduction in uncertainty (e.g. about one's exposure to health risks), social solidarity, social or national iconography (e.g. when a health care system represents the "kind of people we are"), support for manufacturing and innovation in supply chain industries, and even provision of ineffective but popularly demanded treatments (e.g. by traditional healers, 'alternative' medicine, religion-driven interventions). Evidently not all of these objectives are in direct conflict with the enhancement of population health. Neither do they all have equal merit.

What counts as 'cost-effective' that this can be so? Tanzania seems to apply a threshold of acceptability that is even higher than the range recommended by the WHO. The WHO deems an intervention offering a unit of health gain (usually a DALY averted) at a cost under three times GDP per capita ($7,300 in Tanzania) to be 'relatively cost-effective' and one with a cost per unit of outcome less than GDP per capita ($2,400 per capita in Tanzania) to be 'highly cost-effective'. These are already high, even though in Tanzania they would exclude Avastin if the WHO guidance were to be followed. A currently available estimate of a plausible range for the threshold in Tanzania is $45-$912 (Woods 2015). The serious (as distinct from tokenistic) adoption of these high thresholds would rapidly exhaust Tanzania's health budgets, both private and public, and leave many much more cost-effective interventions, with much more impact on the country's health status per dollar, unfunded.

Why it is so wrong for countries to adopt (mostly implicitly) thresholds that acknowledge too many interventions as being worth having? Why is it wrong for the WHO to recommend very high cost-effectiveness thresholds as criteria for selecting interventions in public insurance systems, especially in poor countries?[2] And how might one set about selecting a responsible threshold?

The bookshelf

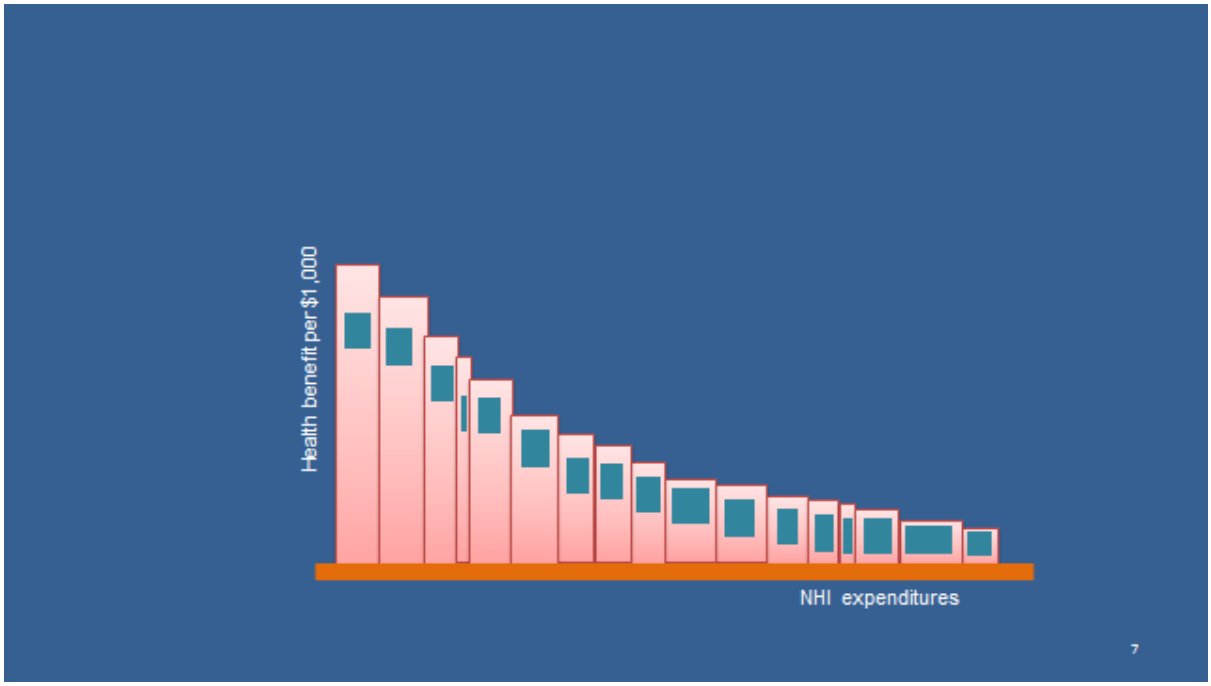

Figure 1 The range of interventions

Imagine a bookshelf such as that in Figure 1 – a very long bookshelf – of health care interventions, each like a book, and ranked according to its effectiveness (its height), with the most effective on the left and the less effective stretching away on the right. Some interventions are disease specific like the cancer treatments just mentioned; some are not disease specific, like interventions to improve childhood nutrition; while others, like community clinics or community health workers, are general delivery platforms or common generic resources available for many diseases and interventions. In all cases, however, we need some acceptable common measure or indicator of the contribution that each intervention makes to health. It must be common, like change in mortality or longevity, or QALYs or DALYs, in order for decision makers to be able to make comparisons of the productivity of

[2] Or even as targets that rich countries have an asserted duty to enable them to implement (a duty unlikely to be recognised by many countries).

each. If they cannot make reasonable comparisons they can hardly make reasonable choices. This may seem a self-evident point. However, nearly all (or at any rate a very large number of) the studies of the effectiveness of interventions for health in LMICs have measures of outcome that ensure comparisons *cannot* be made.

The fatness of each book represents the estimated cost of providing it. This is a combination of the costs of a specific technology, such as a drug, the costs of associated procedures (other medicines, diagnostic services, community services, etc.) for as long as the treatment continues, and the estimated number of people using the intervention in question. A population health promoter will select the first book on the left and add books (that is, further interventions) moving along the shelf until the agency runs out of the money the government has allocated to health, or subscribers have subscribed at current insurance premium rates. At that point all the interventions selected will be effective and only the *most effective* of those that are effective will have been selected. The only services offered under the NHI plan are those to the left. The least cost-effective intervention that is in the plan indicates a threshold of t_o.

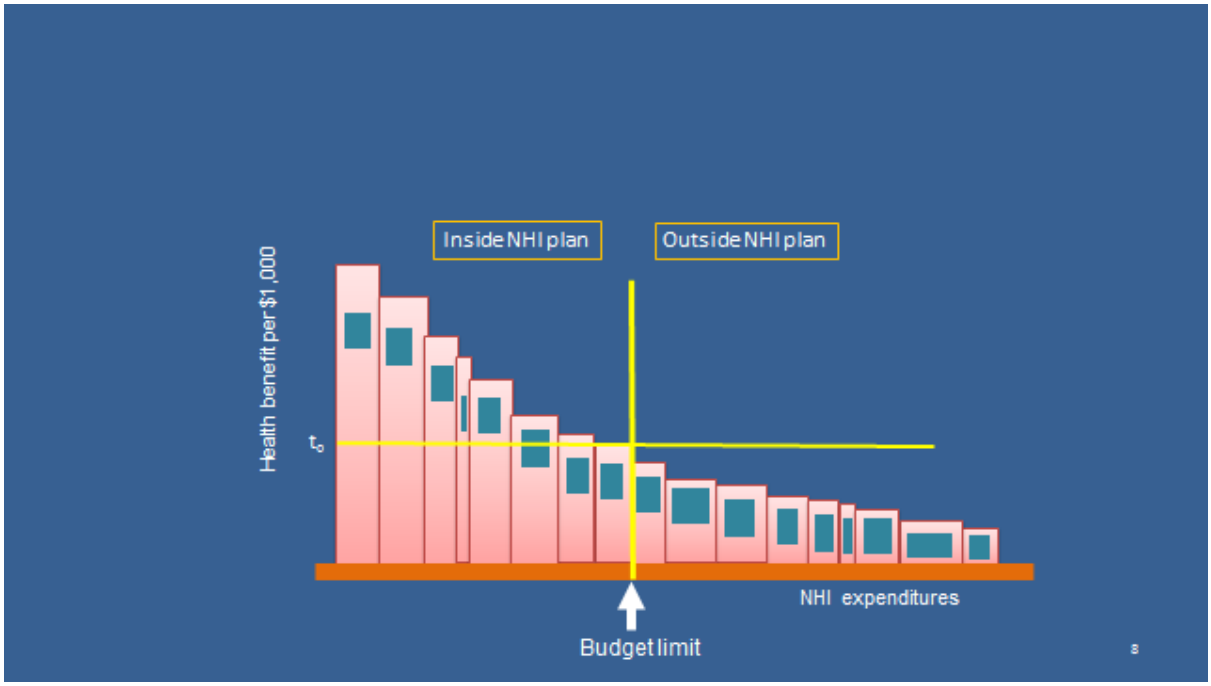

Figure 2 The budget and the threshold

Higher productivity per dollar is equivalent to lower cost per unit of health outcome. The numerator here is health or health gain and the denominator cost. The heights of the books on the shelf, in other words, are the reciprocals of the familiar cost-effectiveness ratios in which the numerator is cost and the denominator is health or health gain.

Why are all effective interventions not in the insured bundle?

The reason why the interventions on the right are not included is *not* because they are ineffective. On the contrary, they are all effective. One would have to go a long way to the right before hitting zero productivity or slipping into the zone of iatrogenesis. The trouble with them is that *they are not effective enough*. If the benchmark test for inclusion of further interventions is the cost-effectiveness of the least cost-effective intervention that is included in the plan, t_o, then they are not *cost*-effective enough either. It immediately follows that merely to demonstrate the effectiveness of an intervention is not enough to ensure its inclusion in the insured bundle. From the viewpoint of

population health there are simply better ways of using the budget. Of course, were the budget to be increased, then further interventions could be added, but this takes us into the realm of 'meso' cost-effectiveness, in which we have to make a judgment about the costs to other public programmes or to private consumption as resources were switched to health and a further judgment is then required as to their value *where they are* relative to the value of the expected increase in health if they were switched[3]. Public health advocates, to be effective, need therefore to demonstrate *relative* effectiveness. One way of doing this is to make direct comparisons between interventions, such as comparing alternative treatments for cancer or for macular degeneration. A less cumbersome procedure is to use the *threshold*, and make comparisons with that.[4]

The threshold and the budget are intimately linked

The determinants of the threshold are fundamentally three: the underlying demographics and disease burden, which affects the productivity of interventions (if that rises, the threshold *t* rises); local environments, customs and values (which affect health and commercial behaviours and the very notion of 'health' and hence again the productivity of interventions); and the budget (if that were to fall, *t* rises). In summary, the threshold depends on (a) the productivity of interventions and (b) the size of the budget. We can see this in the figures. First consider a fall in the budget, ceteris paribus. The budget line moves to the left in Figure 3 and the threshold rises to t_1. Now there are fewer types of intervention in the insured bundle.

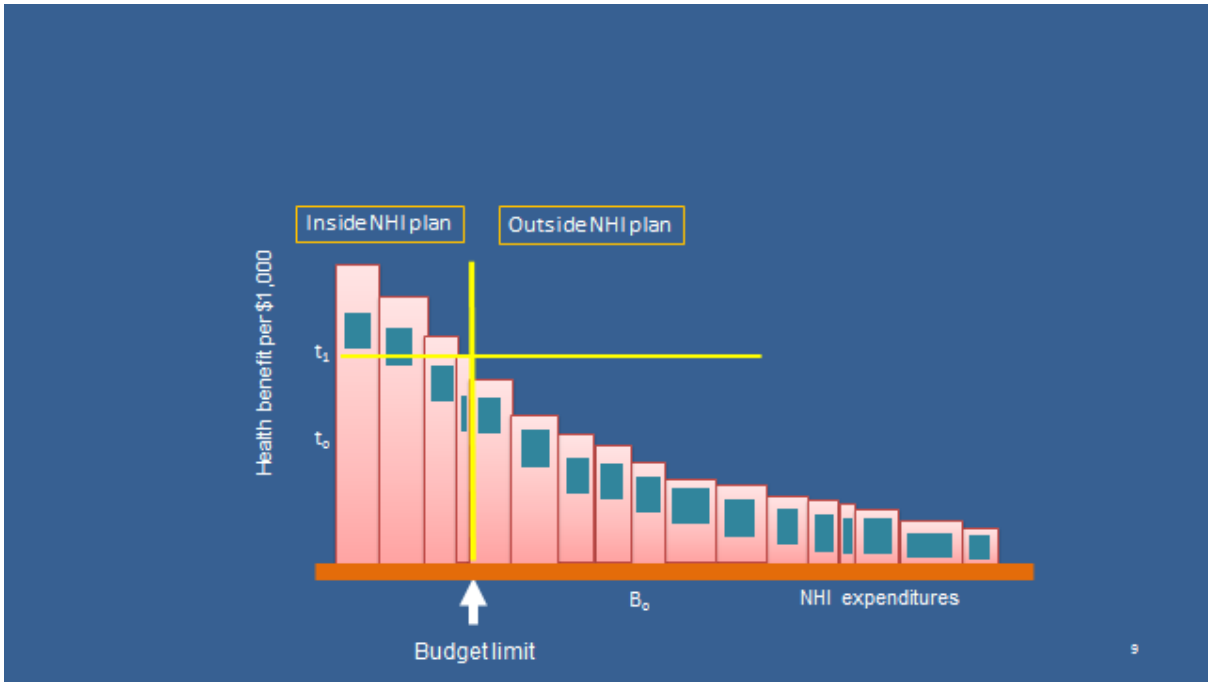

Figure 3 Budget falls and threshold rises

Or let the budget rise. The vertical budget line moves to the right in Figure 4 and the threshold falls to t_2. The variety of interventions rises.

[3] A more general set of sectoral thresholds is implied with effectiveness defined in terms of a more general indicator of well-being, such as "utility" (an indicator of social or collective strength of preference).
[4] We abstract, as previously stated, from the effectiveness of interventions of all kinds on the non-health objectives of health care systems and their budgets.

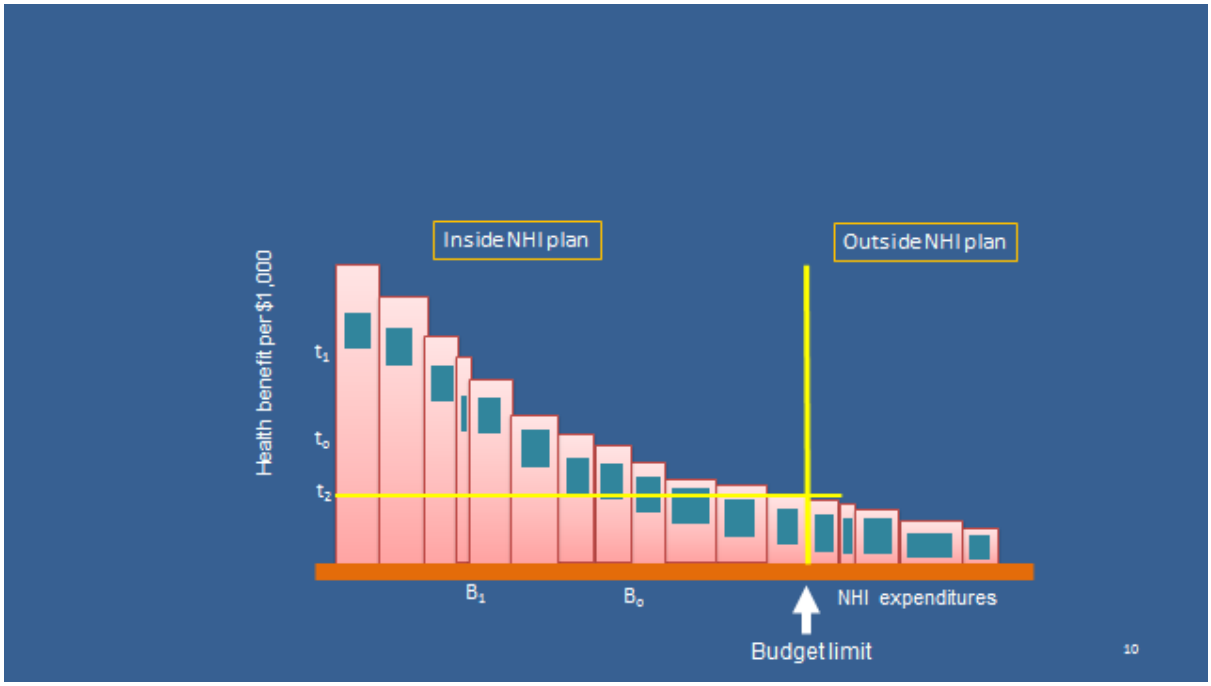

Figure 4 Budget rises and threshold falls

The threshold is not determined only by per capita GDP, which is but one of the determinants, nor is it appropriate to link it in linear fashion to GDP per capita. However, since health care spending has a positive income elasticity, rising incomes enable proportionately higher health budgets and countries with higher incomes to have proportionately higher health spending (see Woods et al., 2015, for some multi-country estimates). One way of looking at the threshold is nonetheless as a demand concept – an implication of a collective willingness to pay for health as expressed by the size of the health budget. This can be contrasted with a supply-side view which is mentioned later.

Choosing badly kills

The morality of proper use of a threshold comes from its impact on people's health, which we take as having a moral worth that usually trumps that of non-health objectives of health care systems. If interventions on the right of the threshold are allowed to replace any on its left, population health falls. In Figure 5 books on the shelf have been swapped from either side of the budget line. The hatched area is the loss of life and/or quality of life from having the wrong things in the plan. Decision makers are typically ignorant as to whether they have the right things assigned to either side of the vertical budget line (Eckermann and Pekarsky 2014) but so long as they always replace interventions having lower productivity per dollar with ones that have higher productivity per dollar, they will be moving in the right direction and, if they also ensure that those that are included have a productivity per dollar that is higher than t_o, then they can be confident of extracting even more health from their health dollars.

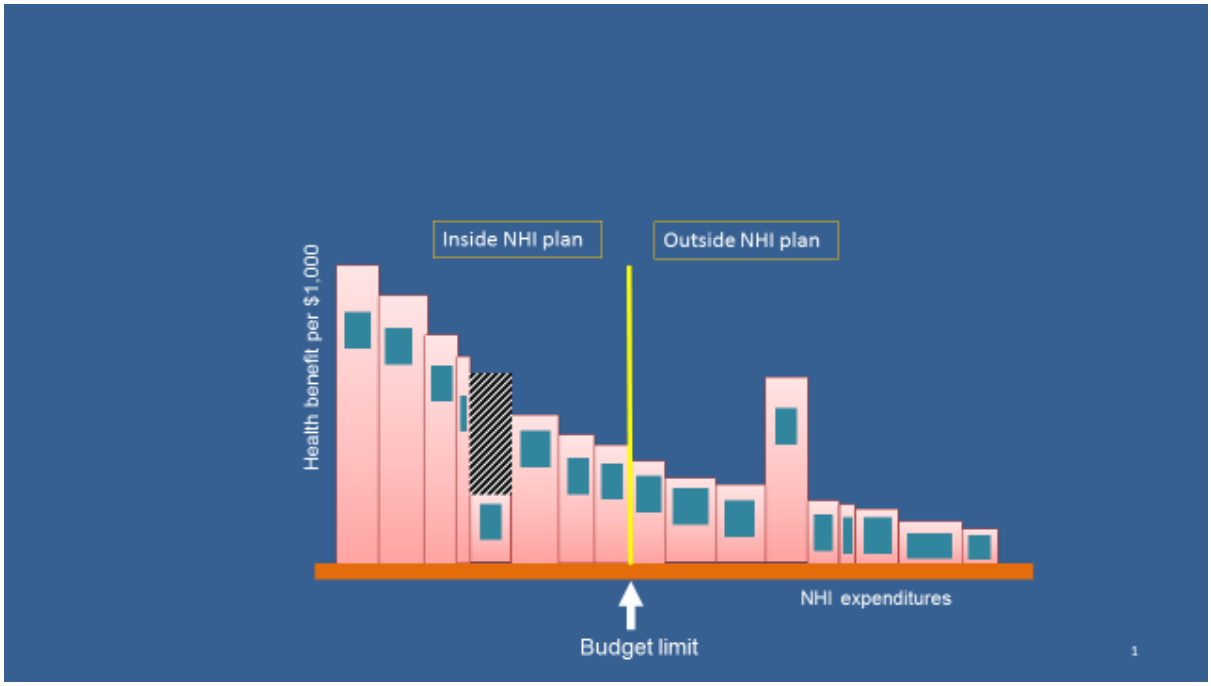

Figure 5 Health loss from poor technology selection

There is the converse: if the low productivity intervention is already in the bundle, then the hatched area represents the health gain from eliminating it and replacing it with the more productive technology on the right. Note the politically difficult and somewhat counterintuitive fact: disinvestment, even in effective technologies, can *increase* health.

First best and second best

The analysis has been purposefully simplified. It has been assumed that all decision makers want is to have as great an impact on population health as possible. We have assumed that decisions are taken in and apply for a single accounting period. Another assumption is that each intervention is not *internally* ranked, as when a procedure is more effective for some types of patient than others, so that some applications of it may be high while others were low in the ranking – and some of them perhaps even lie on the other side of the borderline. Yet another simplification is that the measure of effectiveness is indifferent to the characteristics of the people who gain or lose: an extra year of life or an extra QALY is of equal value whoever gets it. These are all assumptions that can be addressed in specific situations and modified as necessary.

A simplification having deeper consequences is that all the interventions on the left are more productive than all those on the right. If decision makers are planning ahead for a public health insurance plan that is yet to be established using consistent health promoting principles and there is no historical encumbrance of bad past decisions, then they can have some confidence that the interventions selected are more cost-effective than those left out. In this case the assumption is not only descriptively idealistic, it is also realistic. It represents a 'first best' situation. It is a commonplace of welfare economics, however, that the resource allocation 'rules' derived under such circumstances may not be appropriate when these assumptions do not apply (Lipsey and Lancaster 1956). In such cases the theory of 'second best' is required. If now the starting point is the historical inheritance of a set of insured interventions whose evidential base was poor or left unexplored, many of which were selected for reasons other than a plausibly demonstrated highly effective impact on population health, then it is evidently more likely for the insured set to include procedures less effective per dollar than some of those excluded, so less health is lost per dollar

when they are displaced. The estimated (second best) threshold under these circumstances will be lower (the threshold incremental cost-effectiveness ratio will be higher). Put another way, introducing a new intervention into an already established health care package entails, given a constant budget, a greater chance of disinvesting in low productivity interventions, so the real opportunity cost of new interventions is lower. Put more generally, the more internally efficient the health system at any given budget, the higher the opportunity cost of additions to the insured bundle.

The empirical work of Claxton and colleagues (2015) does not make the first best simplifying assumptions but estimates the displacement of interventions, when new ones come in and the budget remains constant, in terms of the *actual* interventions that are dispensed with regardless of their relative productivity. There is no guarantee, for example, that those displaced are the least productive interventions. They may be simply those that are managerially the most convenient in the short run to remove or reduce. However, they indicate the health loss associated with the introduction of a new technology, or its opportunity cost in terms of health. To the extent that this opportunity cost is lower than the loss of the most marginal technology in a first best allocation of the same budget, the second best threshold will appear lower than the first best threshold (the incremental cost-effectiveness ratio will be higher (Eckermann and Pekarsky 2014)). This is a supply-side and behavioural approach to the threshold, which gives an estimate that equals the first best threshold only when the system is in 'equilibrium' at a health-maximising optimum, and the least cost-effective technologies are relatively easy to identify. The method is likely to yield an estimate of opportunity cost (health lost) lower than under the first best allocation of the same budget – how much of an underestimate will depend on the ability of health service commissioners (purchasers) and managers to identify the least-cost technologies that lie within their discretion to eliminate. More importantly, the second best threshold is the appropriate one to use in decisions about revising the content of an established insured bundle since it embodies an estimate of the actual opportunity cost. A specific value of this approach is that, data permitting, it enables decision makers to identify with some degree of precision which interventions may be eliminated at least cost in terms of their contribution to population health.

The implication that the second best cost-effectiveness threshold will typically be higher than the first best threshold should not obscure the fact that an *actual* cost-effectiveness threshold in use may be too high. For example, Claxton et al. (2014) suggest quite strongly that the actual thresholds used by NICE inn England and Wales are too high.

Orphan diseases and an ethical dilemma

Finally, it has also been assumed that making a maximum impact on health is the *only* objective of NHI or UHC. That is obviously not true. One ought at least to add in equity, or distributive fairness and financial protection, as other criteria. However, these complications would clutter the ability of the simple model to yield insights. In particular, decision makers need to be alert to the implication that any departure from the bookshelf principle *on whatever grounds* costs lives, or at least the quality of lives. Departures need, therefore, a strong ethical case.

Suppose that the reason for swapping the two books in Figure 5 lay in the distributional characteristics of each: perhaps the low productivity intervention is a very costly but not very effective treatment for an orphan disease. One has natural humanitarian empathy with patients suffering from diseases like cystic fibrosis, muscular dystrophy, Gaucher's disease, Huntingdon's disease, Hunter's Syndrome or Pompe's disease. But decision makers need to be clear that in replacing a more productive intervention with the orphan treatment, they are causing others to lose lives – or the quality of lives, or both. That may be an acceptable trade-off but a trade-off it is and

one ought not to imagine that attending to other health policy priorities than having the maximum impact on health comes cheap. It is often thought that it is humanitarian to support expensive but not very effective interventions for people with orphan diseases – but it seems not at all humanitarian if to do so mindlessly ignores the health losses imposed on others. This is not to suggest that decision makers ought to lack sympathy for hard cases, only that they should not ditch logic in exercising their sympathy. What is especially tricky about such cases is also that the identity, or at least the characteristics, of a proposed favoured minority group (the orphan disease victims) is known, whereas that of those who lose, whose services are no longer provided from the fixed budget, is not. They are usually anonymous – and easily overlooked. So a further ethical question thus arises – are costs falling on invisible people to be ignored or given a lower weight by virtue of their invisibility? They may even be people with whom decision makers would have no less sympathy than the sympathy they have for those with orphan diseases.

The threshold in low- and middle-income countries

One of the sad truths about health and health care in LMICs is that policy makers are constantly bombarded with claims – many from health economists and public health physicians – for the inclusion of interventions whose only virtue is that they are effective. Childhood interventions, particularly vaccinations, often provide greatest value. In sub-Saharan Africa, for example, rotavirus vaccination has been associated with a cost-per-DALY-averted of $43 (Atherly et al. 2009) and treatment of severe malnutrition costs $53 per DALY-averted (Bachmann 2010). Even within HIV, prevention of mother-to-child HIV transmission costs below $150 per DALY-averted using available interventions (Shah et al. 2011). All of these are likely to be cost-effective choices for countries like Tanzania, yet a large coverage gap remains across all low- and middle-income countries (UNAIDS 2013, Revill et al. 2015). These high-impact and highly cost-effective interventions are seriously under-provided, while advocates routinely make recommendations on the basis of excessively high cost-effectiveness thresholds, or none at all, aiming to promote access in poor communities to new and more expensive therapies with cost-effectiveness ratios 100 times worse than these (Chisholm et al. 2012, Ortegon et al. 2012).

What, then, are the consequences of trying to apply such thresholds?

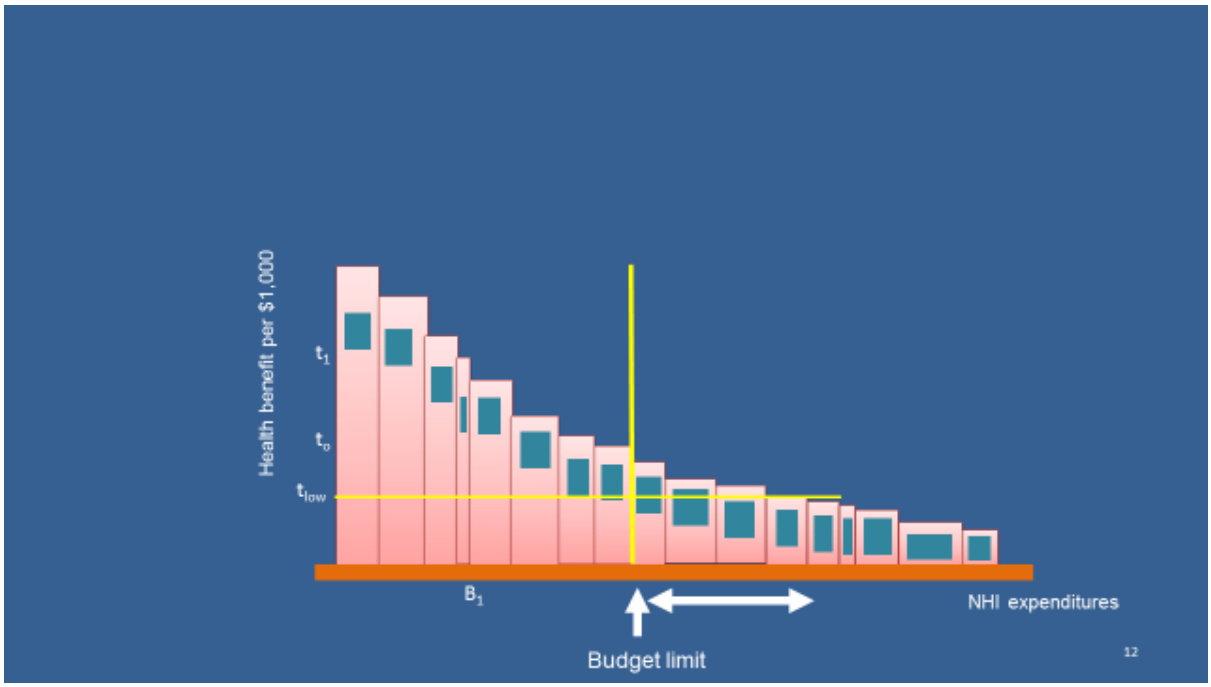

Figure 6 Threshold too low for the given budget

A cost-effectiveness threshold that is too high is a health gain per dollar threshold that is too low. We can see what is likely to happen by returning to the bookshelf. In Figure 6 there is the same array of books as before, for which the threshold was t_o. Setting the threshold at t_{low} is appropriate for a health budget much larger than the one in the figure. It will admit into the insured bundle the additional interventions shown by the arrows. What will happen? In the absence of a further, more rational criterion (e.g. a supplementary threshold of t_o), the outcome is likely to be an arbitrary set of interventions. In an extreme case, the most productive interventions will be replaced by the least productive ones. That is shown by the hatched area in Figure 7. An arbitrary selection of the threshold, which is typically too high, loses lives.

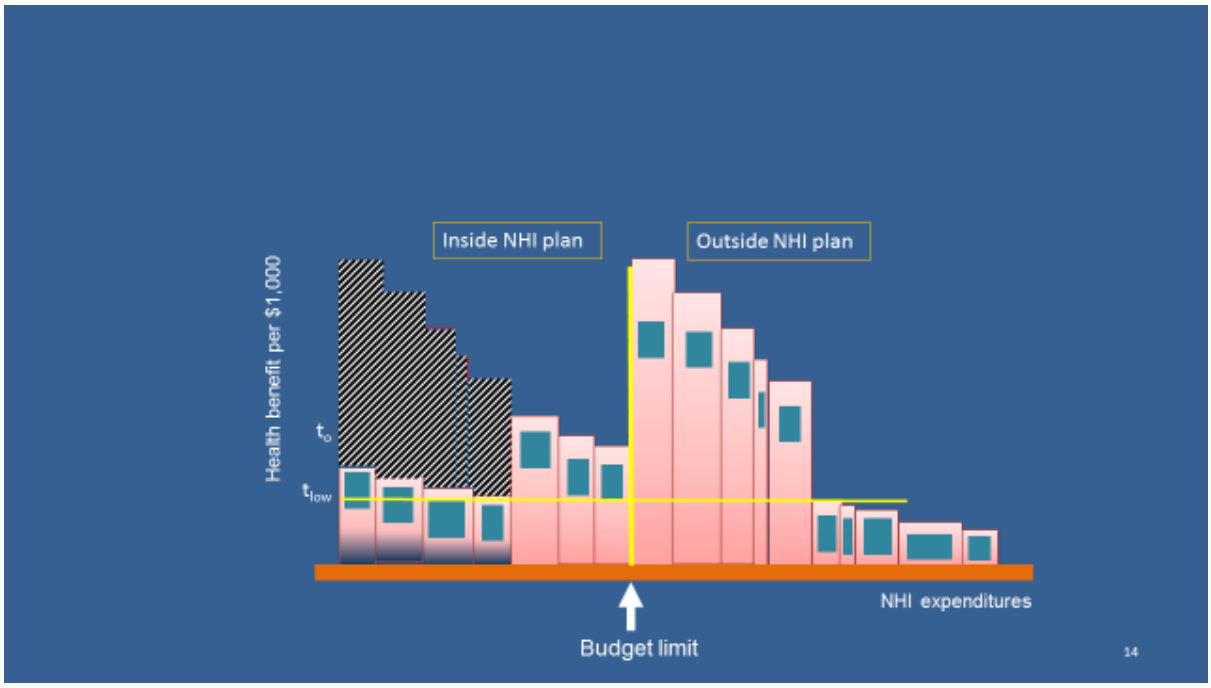

Figure 7 Extreme health loss from threshold set too low

The contrary phenomenon will occur when the threshold is set too high for the budget. This is shown in Figure 8. With t_{high} and the budget as before, all the technologies indicated by the arrows are omitted from the insured package with the consequential loss of life and quality of life shown by the hatched area. The offence to health is compounded by the needless holding back of available funding.

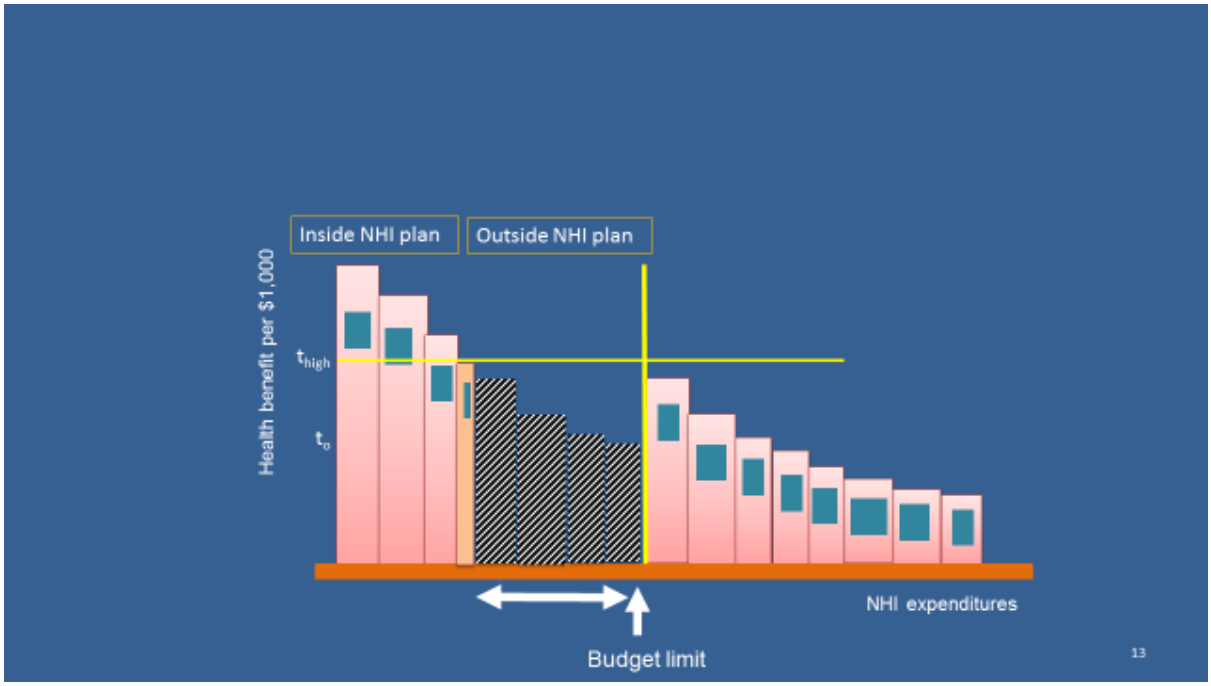

Figure 8 Health loss from threshold too high

Asymmetries in investing and disinvesting

One ought not to underestimate the problems, mainly political, of pressures to invest in interventions that are too costly and that drive out those that are more cost-effective. An intervention that is adopted generates incomes for its manufacturers, its prescribers and is usually gratefully received by its patients. It has a massive interest already vested in it. The same is less true of those that have yet to be approved, which are in direct competition with others yet to be approved, where all the gains are as yet prospective and not as yet vested. It nonetheless remains that the threshold is what marks the boundary between the more and the less cost-effective treatments, given a particular planned rate of expenditure on health care. Any other threshold is arbitrary and harmful to health, notwithstanding the interests (doubtless very vocal) that may be vested in it.

The 'second best' threshold is, however, indisputably hard to estimate, even approximately. Unfortunately, it is also tremendously easy to propose *aspirationa*l (Revill et al. 2015) thresholds that are far too high. These thresholds all implicitly assume that the fraction of GDP that should go on health and health care is much larger than it actually is. They are an implicit form of advocacy. The problem with cost-effectiveness thresholds that are too high is that they define as cost-effective, or even highly cost-effective, treatments that, if implemented, would more than exhaust the available budget and crowd out treatments that deliver more health gain per dollar. At best they can provide a kind of 'long list' of interventions for further consideration. One therefore needs a more realistic supplementary cost-effectiveness threshold to select from the long list: one that is realistic given the circumstances and budget of the country in question. But then why bother with the long list in the first place? The basic truth is that in setting the budget in any country, decision makers thereby also set the threshold, or in setting the threshold they imply the budget. They ought never to set the one without realizing that in so doing they imply the other.

Multiple thresholds?

Some jurisdictions may in the short term have more than one threshold (Danzon et al. 2015b). For example, South Africa has in essence two parallel systems: the private insurance/private provision sector and the public insurance/public provision one. In the short term it makes sense to set a cost-effectiveness threshold for the public sector that is lower than the one implicitly or explicitly set for the private sector. The main challenge then becomes how best to manage the harmonisation of the two over time. (A further significant challenge is how to explain the rationale for dual thresholds to the public and professionals). Tanzania might prudently set a specific threshold for its 'essential' drugs list while it searches for a more general threshold for system-wide application, and a method again of harmonizing the two over time. Some LMICs could have multiple thresholds to reflect the reality that donor funding, for example, has focused on specific disease areas like malaria or HIV/AIDS, where the marginal productivity of disease-specific interventions per dollar has fallen below that of other forms of intervention (say, nutrition). The point in having multiple thresholds is to avoid, not to perpetuate, investments that are not having the greatest possible impact on population health, by identifying where the best pay-offs lie and to signal the necessary shifts in resource allocation.

Differential weights on outcomes and the threshold

The standard application of the principle of horizontal justice requires that people who are alike in some ethically relevant sense be treated alike. There has arisen a concern in some quarters for departing from the usual assumption that a 'QALY is a QALY is a QALY' in favour of weighting QALYs received by those near the end of life more heavily (Paulden et al. 2014, Claxton et al. 2015, McCabe et al. 2015). Making such adjustments, whether on these grounds or any other, has implications for the threshold itself. Paradoxically, increasing the weight attached to particular technologies in order for them to pass the threshold test can also increase the threshold itself. Take the end-of-life case. A candidate technology is particularly to be used for patients near the end of their lives. Supposing the closeness to death to be the only ethically relevant differentiating feature, then ethical consistency requires that all end-of-life cases receive similar weights, including those receiving treatments that are already in the insured bundle. Suppose further that the weighting accorded the candidate technology is sufficient for it to fall below the existing cost-effectiveness threshold. The impact on the threshold depends upon the distribution of end-of-life patients across the technologies in the insured bundle. If they are users of the least productive (the marginal) technology in the bundle, then its effective (weighted) productivity also increases and the cost-effectiveness threshold therefore falls. If it increases sufficiently relative to the next most productive technology, that technology may become the marginal treatment, to be displaced in whole or in part by the new technology. If, on the other hand, such patients are distributed across the inframarginal technologies in the bundle, then the marginal technology remains unchanged and so does the threshold. The candidate technology displaces the existing marginal technology.

The point is that, whatever the merits or otherwise of favouring such groups in considering the introduction of new interventions, similar groups exist elsewhere within the same health care system and are actually receiving current treatments for other conditions. The consequences for them as they are denied services will need normally to be taken into account on grounds of ethical consistency. A judgment has therefore to be reached as to the distribution of such patients across existing insured treatments and the consequential quantitative impact of the differential weights on the threshold.

What is *my* threshold?

Failure to set a threshold, whether or not it is publicly declared, can have similar consequences to setting one that systematically admits too many low productivity interventions. If jurisdictions deliberately fail to use a threshold they should probably stop pretending that that they are trying to have the maximum impact on their people's health. Many countries are shy about being explicit about thresholds (Canada, the USA). Federal structures are easily capable of permitting the simultaneous existence of multiple thresholds (one for each province or state or public programme), mostly implicit rather than explicit. All are ways of ensuring that population health is not maximised.

One way of avoiding setting thresholds aspirationally is to 'threshold search' (Culyer et al 2007) by identifying the least cost-effective intervention currently provided and the most cost-effective intervention not yet routinely available. This approach might be suited to a two- or multi-threshold country. There are a reasonable number of available economic evaluations that are probably generalizable and applicable in most jurisdictions. One might investigate the cost-effectiveness of interventions falling just inside and just outside the insured bundle in various jurisdictions and triangulate towards a reasonable approximation to 'second best' threshold. Yet another approach, appropriate for countries with very low public expenditures, is to proceed pragmatically with self-evidently cost-effective programmes, with scaling-up determined judgmentally, but evaluating each scale-up and each newly added intervention so that over, say, a five year period sufficient information became available about the cost-effectiveness of the programmes being supported and the pressure on budgets. The cost-effectiveness of the least cost-effective programme being supported then becomes a provisional threshold and new programmes with lower cost-effectiveness would not be recommended.

Another possibility is to conduct low-cost small scale pilot studies of prima facie highly cost-effective interventions which could then be scaled up or not as and when their efficiency is confirmed or disconfirmed. The Thais have been good at this (e.g. Teerawattananon et al. 2009, Teerawattananon et al. 2014).

Yet another approach is to estimate the productivity of health care expenditure *across* countries. Multi-country panel data show that health outcomes improve as countries increase spending on health care, although at a diminishing rate. Understanding this relationship could indicate which interventions are likely to increase or reduce productivity in the health sectors of different jurisdictions with particular levels of resources and healthcare needs.

The most complete approach, where the data exist to implement it, is the supply-side method developed by Claxton et al. (Claxton et al. 2015). This econometric work does not make the simplifying assumption that no technology in use has a lower productivity than the threshold, and exploits the existence of programme budgets in the National Health Service. These cover 23 budgets. Differences in them can be linked to differences in mortality and, with some further assumptions, to QALYs. The central estimate of the threshold using this method for 2008 expenditure and 2008–10 mortality was £12,936 per QALY, well below the threshold range of £20k-30k used by NICE.

Aiding not replacing thought and judgment

However, perfect precision is not in general required. What is needed is an understanding of the meaning of the threshold and some idea of its likely order of magnitude in any given context. The purpose of this information is to inform thought and judgment, not to replace either. The epidemiological science is always contestable, the endpoints of trials are rarely far enough off, what

is demonstrably efficacious may not be effective, the coverage of costs is often incomplete, substantial uncertainty attaches to many key variables, full scientific consensus is rare. The cost-effectiveness of most interventions (inside and outside an insured bundle) remains conjectural. The evidence therefore never speaks for itself. In any case, judging what to include in the insured bundle cannot be solely based on evidence, even if it is good evidence (Culyer 2014) because a decision also depends upon the objectives set for the health care system and on the value judgments embodied in the criteria for including or excluding treatments in or from the insured bundle (Rawlins and Culyer 2004). But aids to better judgment are valuable, provided they are understood and the science supporting them is honest science. They are valuable, not only because they increase the chance that good decisions will be made but also because they nearly always involve the participation of others than 'experts', thereby gaining public credibility, and because the process of exercising one's judgment, provided it is not conducted in secrecy, can be publicly defended and can lead to a public media and a citizenry that also understands and judges in an informed way.

In summary…

We started with the simple idea that more health is a good thing and then showed that to achieve more health

- it must be possible to compare interventions in terms of their impact on a common measure of health;
- mere effectiveness is not a persuasive case for inclusion in public insurance plans;
- public health advocates need to address issues of relative effectiveness if they are to be more effective advocates for public health;
- a 'first best' benchmark or threshold ratio of health gain to expenditure identifies the least effective intervention that should be included in a public insurance plan;
- the reciprocal of this ratio – the cost-effectiveness threshold – will rise or fall as the health budget rises or falls (ceteris paribus);
- setting thresholds too high or too low costs lives;
- failure to set a cost-effectiveness threshold also involves avertable deaths and morbidity;
- the threshold cannot be set independently of the health budget;
- what is cost-effective in one jurisdiction may be cost-ineffective in another (depending on the budget, local prices and costs, disease burden and local values);
- the threshold can be approached from either the demand-side or the supply side;
- the two are equivalent only in a health-maximising equilibrium;
- the supply-side approach generates an estimate of a 'second best' threshold that is higher than the 'first best';
- the 'second best' threshold is the one generally to be preferred in decisions about adding or subtracting interventions in an established public insurance package;
- multiple thresholds are implied by systems having distinct and separable health budgets;
- disinvestment involves eliminating *effective* technologies from the insured bundle;
- differential (positive) weighting of beneficiaries' health gains may increase the threshold;
- anonymity and identity are factors that may affect the interpretation of the threshold.

Finally, and implicitly throughout, the true opportunity cost of health care in a community, where the effectiveness of interventions is determined by their impact on health, is not to be measured in money – but in health itself.

References

Atherly D., R. Dreibelbis U.D. Parashar, C. Levin, J.Wecker, R.D. Rheingans (2009), "Rotavirus vaccination: cost-effectiveness and impact on child mortality in developing countries". *The Journal of Infectious Diseases.*200 Suppl 1: S28-S38.

Bachmann, M.O. (2010), "Cost-effectiveness of community-based treatment of severe acute malnutrition in children". *Expert Review of Pharmacoeconomics & Outcomes Research.* 10: 605-612.

Birch, S., Gafni, A. (1992) "Cost effectiveness/utility analysis: do current decision rules lead us to where we want to be?" *Journal of Health Economics*, 11: 279–96.

Chambers, J.D., Neumann, P.J., Buxton, M.J. (2010) "Does Medicare have an implicit cost-effectiveness threshold?" *Medical Decision Making*, 30: E14-27.

Chisholm D., R. Baltussen R, D.B. Evans, et al. (2012), "What are the priorities for prevention and control of non-communicable diseases and injuries in sub-Saharan Africa and South East Asia?" *British Medical Journal,* 344: e586.

Claxton, K., S. Martin, M. Soares, N. Rice, E. Spackman, S. Hinde, N. Devlin, P.C. Smith, and M. Sculpher (2015), "Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold", *Health Technology Assessment,* 19, DOI: 10.3310/hta19140.

Claxton, K., M. Sculpher, S. Palmer and A.J. Culyer (2015) "Cause for concern: is NICE failing to uphold its responsibilities to all NHS patients?" *Health Economics*, 24: 1–7.

Culyer, A.J. (2014), "Are there *really* ten good arguments for a societal perspective in the economic evaluations of medical innovations?" in A J Culyer and G Kobelt (eds.) *Portrait of a Health Economist: Festschrift in Honour of Bengt Jönsson*, Lund: Institute of Health Economics, 2014, 33-38.

Culyer, A.J. (2015) "Why do/should we do economic evaluation?" *Value in Health Spotlight*, 1: 8-10.

Culyer, A.J., C. McCabe, A. H. Briggs, K. Claxton, M. Buxton, R. L. Akehurst, M. Sculpher, J. Brazier. (2007) "Searching for a threshold, not setting one: the role of the National Institute of Health and Clinical Excellence", *Journal of Health Service Research and Policy,* 12: 56-58.

Danzon, P.M., Towse, A., Mulcahy, A.W. (2015a) "Setting cost-effectiveness thresholds as a means to achieve appropriate drug prices in rich and poor countries", *Health Affairs*, 30: 1529-1538.

Danzon, P.M., Towse, A., Mestre-Ferrandiz, J. (2015b) "Value-based differential pricing: efficient prices for drugs in a global context", *Health Economics*, 24: 294-301.

Devlin, N., Parkin, D. (2004) "Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis", *Health Economics*, 13: 437–52.

Doubilet, P., Weinstein, M.C, McNeil, B.J. (1986) "Use and misuse of the term 'cost effective' in medicine", *New England Journal of Medicine*, 314: 253–56.

Eckermann, S., Pekarsky, B. (2014) "Can the real opportunity cost stand up: displaced services, the straw man outside the room", *PharmacoEconomics*, 32: 319-25.

Gafni A., Birch, S. (2006) “Incremental cost-effectiveness ratios (ICERs): The silence of the lambda”, *Social Science and Medicine*, 62: 2091–2100.

George, B., Harris, A., Mitchell, A. (2001) “Cost-effectiveness analysis and the consistency of decision making”, *PharmacoEconomics*, 19: 1103-1109.

Laupacis, A., Feeny, D., Detsky, A., Tugwell, P. (1992) “How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations”, *Canadian Medical Association Journal*, 146: 473-81.

Lipsey, R.G., Lancaster, K. (1956) “The general theory of second best”, *Review of Economic Studies*, 24, 11-32.

Marseille, E., Larson, B., Kazi, D.S., Kahn, J.K., Rosen, S. (2015) “Thresholds for the cost-effectjveness of interventions: alternative approaches”, *Bulletin of the World Health Organisation*, 93: 118-124.

McCabe, C., Paulden, M., O’Mahony, J.F., Edlin, R. and Culyer, A.J. (2015), “Life at a premium: considering an end-of-life premium in value-based reimbursement”, *Value in Health*, 18, A6-A7.

Newall, A.T., Jit, M., Hutubessy, R. (2014), “Are current cost-effectiveness thresholds for *PharmacoEconomics*, DOI 10.1007/s40273-014-0162-x.

NICE (2013a) *Bevacizumab in combination with paclitaxel and carboplatin for first-line treatment of advanced ovarian cancer*, London, NICE.

NICE (2013b), *Ranibizumab for treating visual impairment caused by macular oedema secondary to retinal vein occlusion*, London, NICE.

Nyhan, B. (2010). "Why the ‘Death Panel’ myth mouldn't die: misinformation in the Health Care Reform debate", *The Forum* (Berkeley Electronic Press) 8 doi:10.2202/1540-8884.1354.

Ortegon M., S. Lim, D. Chisholm, S. Mendis (2012), “Cost effectiveness of strategies to combat cardiovascular disease, diabetes, and tobacco use in sub-Saharan Africa and South East Asia: mathematical modelling study”. *British Medical Journal,* 344: e607.

Paulden, M. J.F. O’Mahony, A.J. Culyer, C. McCabe (2014), “Some inconsistencies in NICE’s consideration of social values”, *PharmacoEconomics*, 32: 1043-53.

Rawlins, M.D. A.J. Culyer (2004) “National Institute for Clinical Excellence and its value judgements”, *British Medical Journal*, 329: 224-227.

Revill, P., S. Walker, J. Madan, A. Ciaranello, T. Mwase, D.M. Gibb, K. Claxton, M.J. Sculpher (2015), *Using cost-effectiveness thresholds to determine value for money in low- and middle-income country healthcare systems: Are current international norms fit for purpose?* CHE Research Paper 98, University of York, Centre for Health Economics.

Shah, M., B. Johns, A. Abimiku, D.G. Walker (2011), “Cost-effectiveness of new WHO recommendations for prevention of mother-to-child transmission of HIV in a resource-limited setting”. *AIDS,* 25: 1093-1102.

Tanzania Ministry of Health and Social Welfare (2013*), Standard Treatment Guidelines and Essential Medicines List*, Dar es Salaam, Ministry of Health and Social Welfare.

Teerawattananon, K., M. Chaw-Yin, K. Wongkittirux, Y. Teerawattananon, B. Chinkulkitnivat, S. Orprayoon, S. Kusaku, S. Tengtrisorn, W. Jenchitr, (2014), "Assessing the accuracy and feasibility of a refractive error screening program conducted by school teachers in pre-primary and primary schools in Thailand", *Plos One*, 9, 6.

Teerawattananon, Y., Y. Leelukhanaveera, P. Hanvoravongchai, M. Thavorncharoensap, L. Ingsrisawang, S. Tantivess, U. Chaikledkaew, N. Hiransuthikul and C. Leartpiriyasuwat (2009), *The potential of provider-initiated voluntary HIV counselling and testing at health care settings in Thailand*, Nonthaburi, HITAP (Health Intervention and Technology Assessment Program).

Towse, A., et al. (eds.) (2002) *Cost-effectiveness thresholds: economic and ethical issues*. London: Office of Health Economics.

UNAIDS (2013), *Global Report: UNAIDS report on the global AIDS epidemic 2013,* New York, United Nations.

Weinstein, M.C., Stason, W. (1977) "Foundations of cost effectiveness analysis for health and medical practices", *New England Journal of Medicine*, 296: 716–21.

Weinstein, M.C., Zeckhauser, R. (1973) "Critical ratios and efficient allocation", *Journal of Public Economics*, 2: 147–57.

Woods, B., P. Revill, M. Sculpher and K. Claxton (2015), *Country-level cost-effectiveness thresholds: initial estimates and the need for further research*, CHE Research Paper 109, University of York, Centre for Health Economics.

World Health Organisation (2003), *The selection and use of essential medicine*s, WHO Technical Report Series, No. 914, Geneva, WHO.