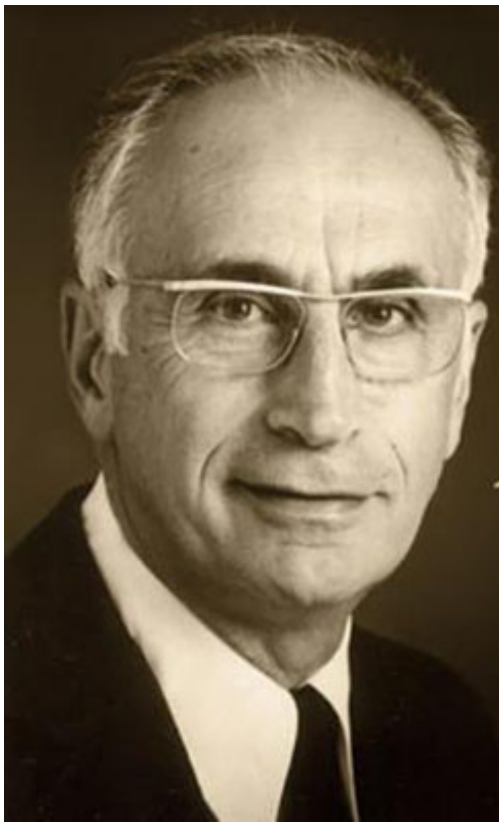
Alchian, Williams, bookshelves and cost-effectiveness thresholds: the power of multum in parvo

The 2015 Hall Lecture (CAHSPR)

Tony Culyer

(IHPME, University of Toronto and CHE University of York)

Two people more than any other have shaped my thinking in economics. One was Armen Alchian and the other Alan Williams. I knew Alchian for only one year of my life – as a graduate student at UCLA in 1964-5. His was a towering presence at UCLA and we were all in awe of him. His teaching style has been nicely described by David Glasner[1]: "Armed with nothing but a chalkboard and piece of chalk, Alchian would lead us relatively painlessly from confusion to clarity, from obscurity to enlightenment. The key concepts with which to approach any problem were to understand the choices available to those involved, to define the relevant costs, and to understand the constraints under which choices are made." His style was Socratic. The logic was relentless. He had an air of amused, philosophical detachment – never condescending but always inviting one as an equal partner to enter his world, the world of the ultimate economist's economist. He would take a topic currently in the news (not necessarily a conventionally 'economic' topic) and question us about it, then using the simplest first principles he would dissect it, explain the phenomenon – always delectable when the explanation was counter-intuitive. A classic example of a simple but counter-intuitive idea is his invention of what has



[1] See his blog at http://uneasymoney.com/2013/02/25/armen-alchian-the-economists-economist/.

become known as the third law of demand: if the prices of two substitutes, such as high and low grades of apples or wine, are both increased by a fixed per-unit amount like a transportation cost, relatively more of the *higher* priced good will be consumed (Alchian & Allen 1963).

Ken Arrow once told me that Alchian was the brightest economics student Stanford ever had. For me he was an inspiration.

Politically, he was a libertarian.

Alan Williams was, by contrast, a lifelong socialist. I met him first in 1960 as my interviewer when I was seeking admission as an undergraduate to Exeter University. He had similar qualities to Alchian – relentless logic that started from the most basic principles: constraints (real and imagined), demand (private or public), opportunity cost, and an explicit normative idea of social welfare. In addition, and for me a big additional attraction, Williams was a great geometrician. His three dimensional depictions on two-dimensional chalk boards were to be marvelled at (Williams 1963). He was by nature a welfare economist. He led me more specifically into health economics and into what I have come to call extra-welfarism. He taught me public finance as an undergraduate at Exeter and we subsequently became colleagues at York. This year is the thirtieth anniversary of a famous article by him in the BMJ (Williams 1985) which I am building upon here. Williams died in 2005 at the young age of 77. Alchian died in 2013 at the age of 98. Their memory deserves to be kept bright.

The point of talking about them today is that they both exemplified in their thinking and teaching the power of simplicity: multum in parvo – much from

little, the casting away of all frills and complications to get to the heart of a puzzle and then to solve it step by step. One might call it relentless reductionism. I cannot hope to equal them in virtuosity but what I want to do here is to try to apply their style of thinking to an important problem.

What treatments should be included in public insurance schemes?

First, some background. The World Health Organisation 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 – to the effect 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. They stand, nonetheless, on Tanzania's list of cost-effective "essential" medicines. 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 World Health Organization. The World Health Organisation deems an intervention

offering a unit of health gain (usually a DALY averted) at a cost under three times GDP per capita ($7,329 in Tanzania) to be “relatively cost-effective” and one with a cost per unit of outcome less than GDP per capita ($2,443 per capita in Tanzania) to be “highly cost-effective”. These are already too high, even though they would exclude Avastin if the WHO guidance were to be followed. The best 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.

The problem is this: 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 World Health Organisation to recommend cost-effectiveness thresholds that are too high as criteria for selecting interventions in public insurance systems, especially in poor countries? This is not the same as the positive question (which Alchian would have asked): “why *did* the WHO make such recommendations?”, but the normative question as to *why* it was wrong – wrong to the point of irresponsibility. And how might one set about selecting a responsible threshold?

The bookshelf of interventions

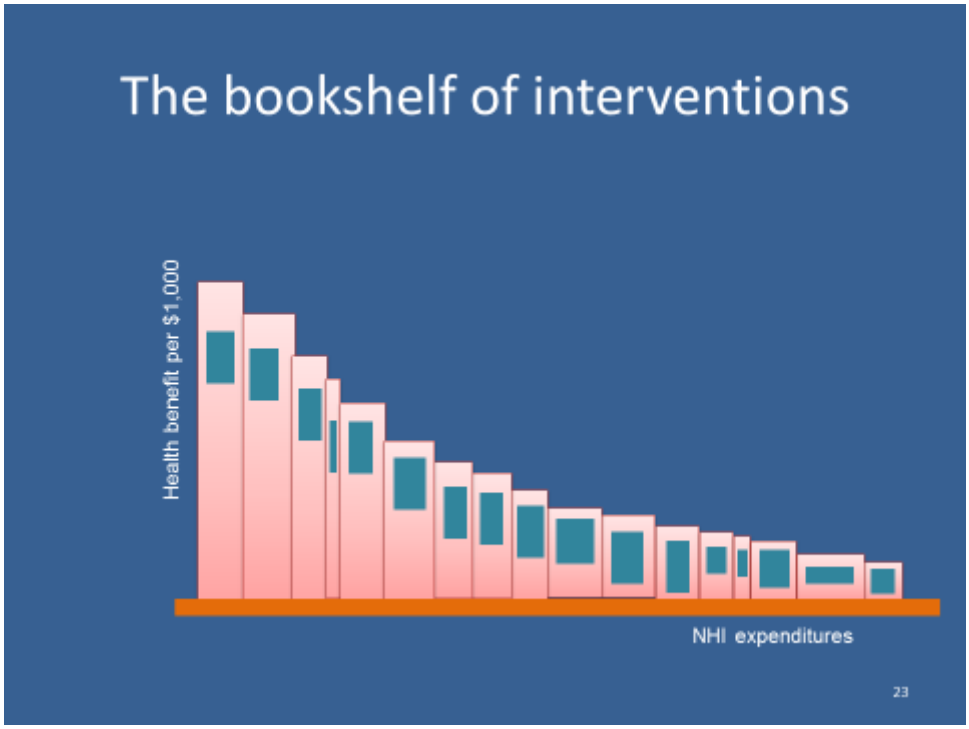


Figure 1 The range of interventions

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. As some of you will doubtless recognise, this is a pictorial representation of a famous so-called league table by Alan Williams (Williams 1985). 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 us to be able to make comparisons of the productivity of each. If we can't make reasonable comparisons we can hardly make reasonable choices. This is my first simple – you might think self-evident – insight. Alas, 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.

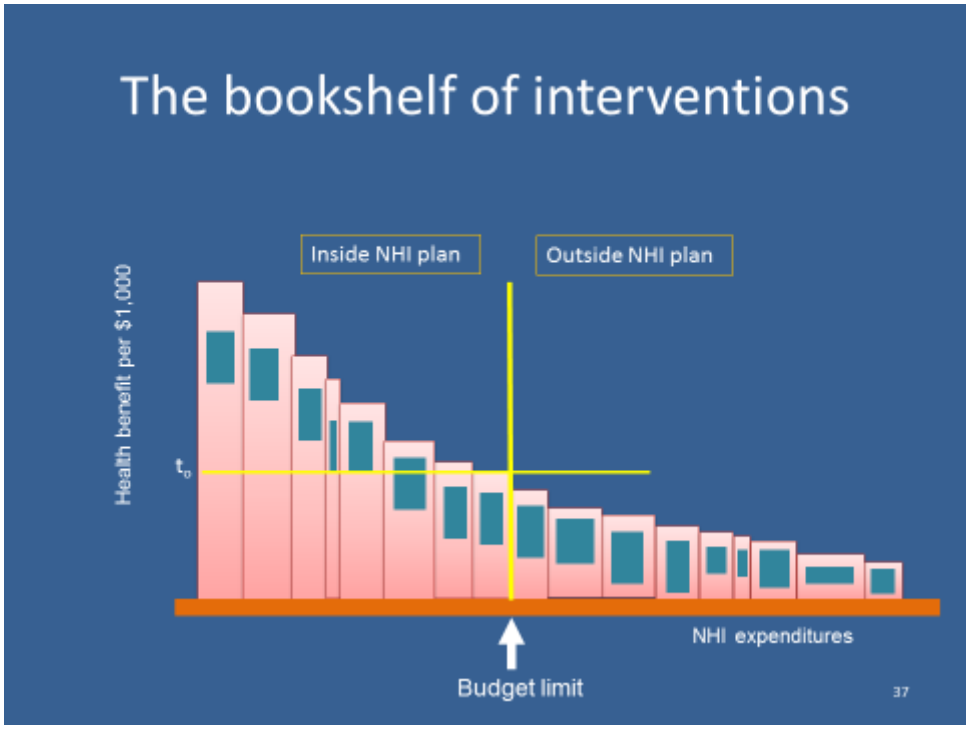


Figure 2 The budget and the threshold

The fatness of each book represents the 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. So, being health maximisers as best we are able (I shall make this assumption), we select the first book on the left and add books (that is, further interventions) moving along the shelf until we run out of the money the government has allocated to health (Figure 2). At that point all the interventions we have selected will be effective and we will also have selected as inside the plan only the *most effective* of those that are effective. The only services we offer under our health maximising NHI plan are those to the left. The least cost-effective intervention that is in the plan indicates a threshold of t_o.

Note that higher productivity per dollar is equivalent to lower cost per unit of health outcome. My numerator 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. We would have to go a long way to the right before we hit zero productivity or even slipped 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 – or ought not to be enough – to ensure its inclusion in the insured bundle. 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 macro 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. Public health advocates, to be effective, need therefore to demonstrate *relative* effectiveness, and this is my third critical insight. 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.

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.

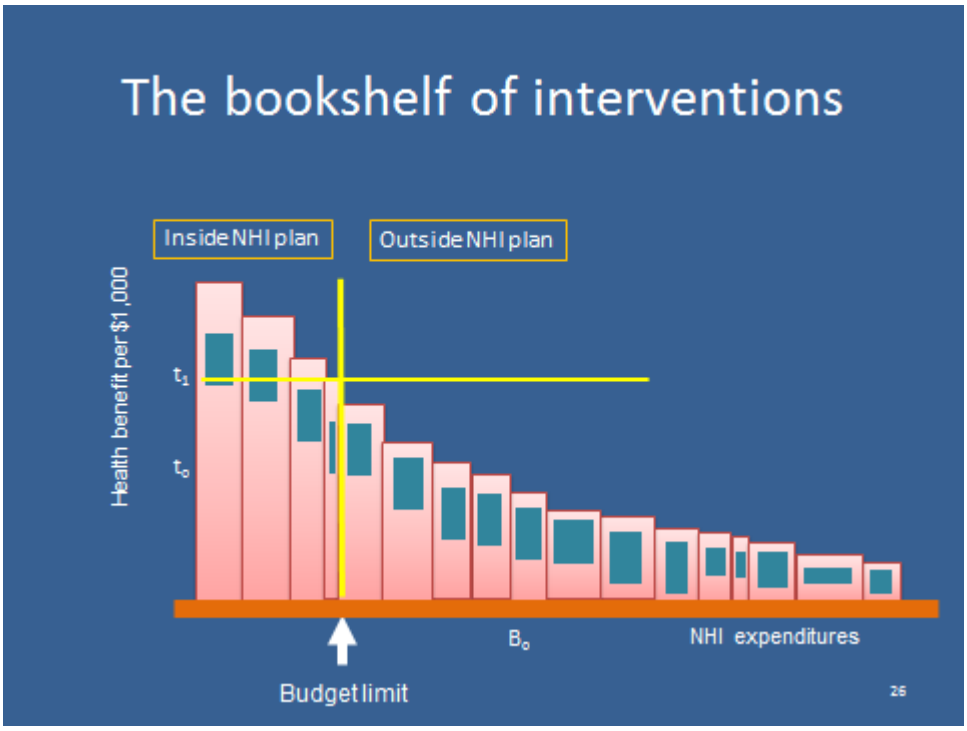


Figure 3 Budget falls and threshold rises

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.

Note that 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. Of course, since health care spending has a positive income elasticity we expect rising incomes to enable proportionately higher health budgets and countries with higher incomes to have proportionately higher health spends. (See Woods et al., 2015, for some reasonable multi-country

estimates). One way of looking at the threshold is as a demand concept – an implication of a collective willingness to pay of a health maximising community as expressed by the size of the health budget. This can be contrasted with a supply-side view which I shall mention later.

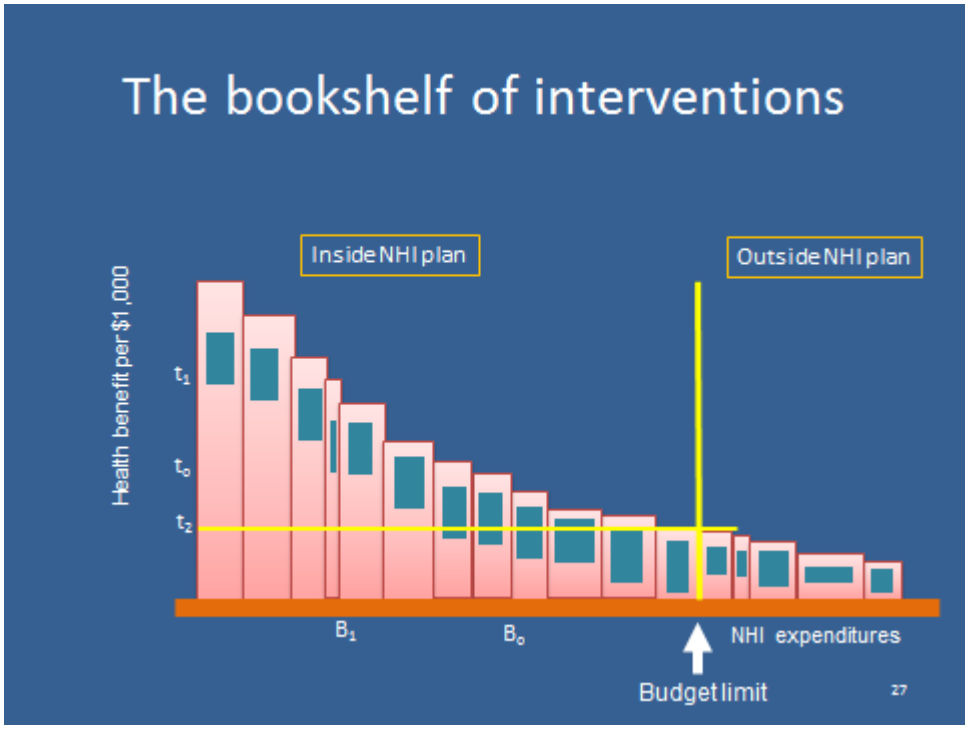


Figure 4 Budget rises and threshold falls

Choosing badly kills

If we allow interventions on the right of the threshold to replace any on its left, we will reduce population health. In Figure 5 I have swapped two books on the shelf from either side of the budget line. The red area is the loss of life and/or quality of life from having the wrong things in the plan. We are, needless to say, typically extremely ignorant as to whether we have the right things assigned to either side of the vertical budget line, but so long as we always replace interventions having lower productivity per dollar with ones that have higher productivity per dollar, we will be moving in the right direction and, if we also ensure that those that are included have a productivity per dollar that is

higher than *t*, then we can be confident of extracting even more health from our health dollars.

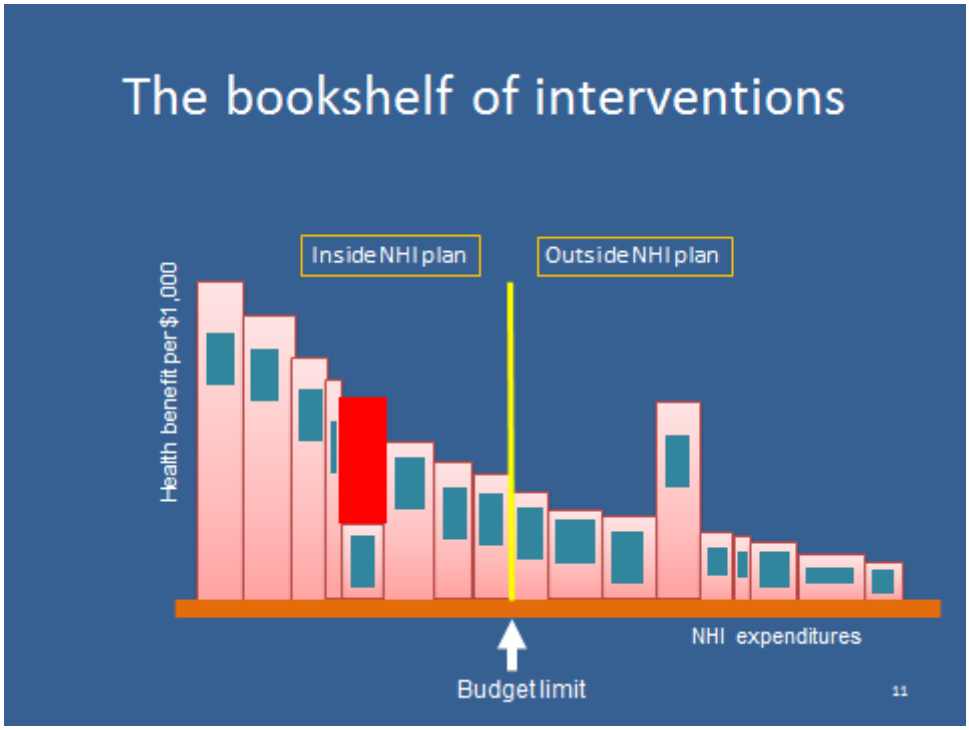


Figure 5 Health loss from poor technology selection

Note the converse: if the low productivity intervention is already in the bundle, then the red 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.

Simplifications

I have simplified. The biggest simplification is probably that of assuming that all the interventions on the left are more productive than all those on the right. If we are planning ahead for a public health insurance plan that is yet to be established using consistent health maximising principles and we have no historical encumbrance of bad past decisions, then we can have some confidence that the interventions selected were more cost-effective than those left out, at least for the time being. In this case my assumption is not only

descriptively idealistic, it is also realistic. If, however, 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 high effectiveness, then it is evidently more likely for the insured set to include procedures less effective per dollar than some of those excluded. The estimated threshold under these circumstances will be lower (the threshold incremental cost-effectiveness ratio will be higher). Put another way, if you want to introduce a new intervention into an already established health care package, and the budget is constant, there is 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.

I note in passing that the famous empirical work of Claxton and colleagues (2015) at York, does not make my simplifying assumption, 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 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 higher than the loss of the truly most marginal technology in use, the threshold will appear higher than the 'true' threshold (the incremental cost-effectiveness ratio will be lower). This is a supply-side and behavioural approach to the threshold, which equals the demand-side threshold only when the system is in "equilibrium", i.e. at a health-maximising optimum and the least cost-effective technologies are relatively easy to identify. The method is likely to yield an underestimate of the 'true' threshold – how much of one 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.

I have also assumed that all we want is to have as great an impact on population health as we can. 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 I have assumed 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.

Orphan diseases and an ethical dilemma

Finally, I have also assumed that making a maximum impact on health is the *only* objective of national health insurance (NHI) or Universal Health Coverage (UHC). That's obviously not true. We ought at least to add in equity, or distributive fairness and financial protection, as other criteria. However, these complications would clutter the ability of our simple model to yield insights. In particular, we need to note that any departure from the bookshelf principle costs lives, or at least the quality of lives. Suppose that the reason for swapping the two books in the figure lay in the distributional characteristics of each. Let's say that 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 we need to be clear that in replacing a more productive intervention with the orphan treatment, we 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 losses imposed on others. I am not suggesting (I hope that is obvious) that we ought to lack sympathy for hard cases, only that we should not ditch logic in exercising our sympathy. What is especially tricky about such cases is also that the identity or at least the characteristics of the favoured minority group (the orphan disease victims) is known, whereas that of those who lose 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 people with whom one would have no less sympathy than the sympathy we have for those with orphan diseases

The threshold and 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 liken 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 absurdly high WHO 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). That sort of advocacy costs lives!

Let us return to the WHO recommendations that a DALY averted at a cost under three times GDP per capita is “relatively cost-effective” and one averted at a cost less than GDP per capita is “highly cost-effective”. For almost every country in the world, regardless of its wealth, these imply cost-effectiveness thresholds that are too high. What, then, are the consequences of trying to apply such thresholds?

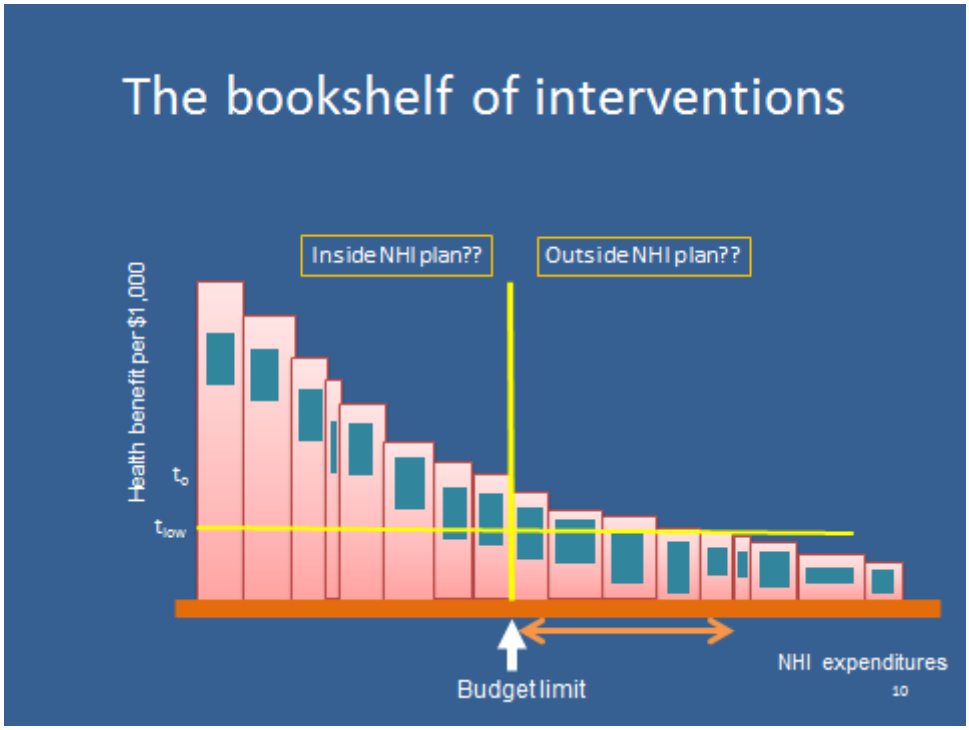


Figure 6 Threshold too low for the given budget

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. Here we have the same array of books as before, for which the threshold was t_o. Following WHO advice, we now set the threshold at t_{low}. This is the threshold that 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 brown 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 red area in the next figure. An arbitrary selection of the threshold, which is typically too high, loses lives.

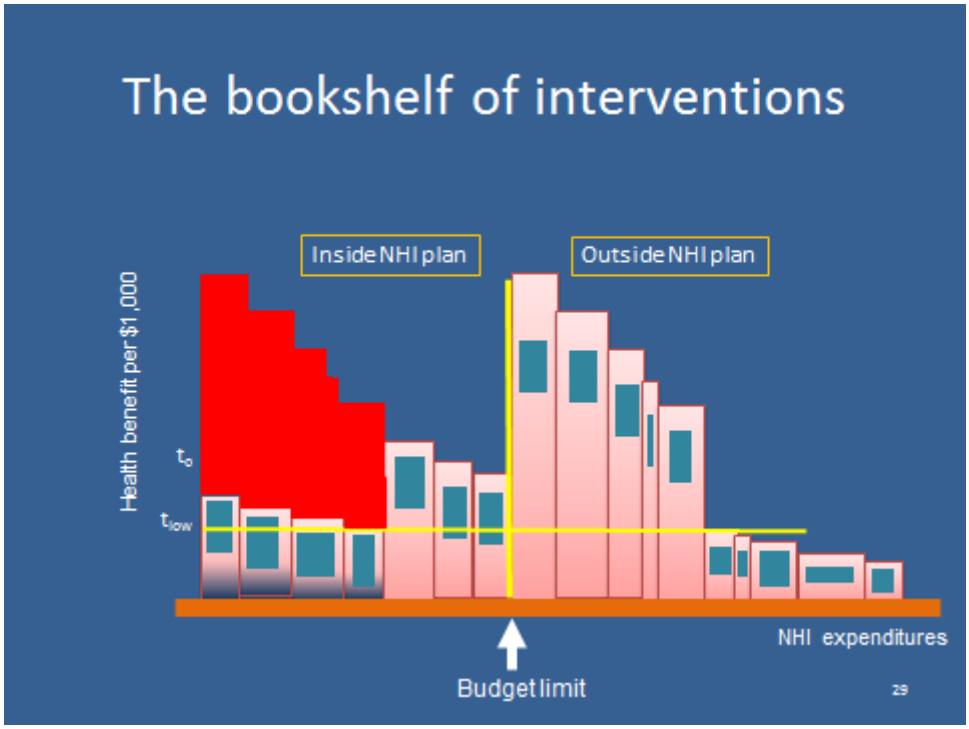


Figure 7 Extreme health loss from threshold too low

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 red shaded area. The offence to health is compounded by the retention of available funding.

Asymmetries in investing and disinvesting

I do not 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

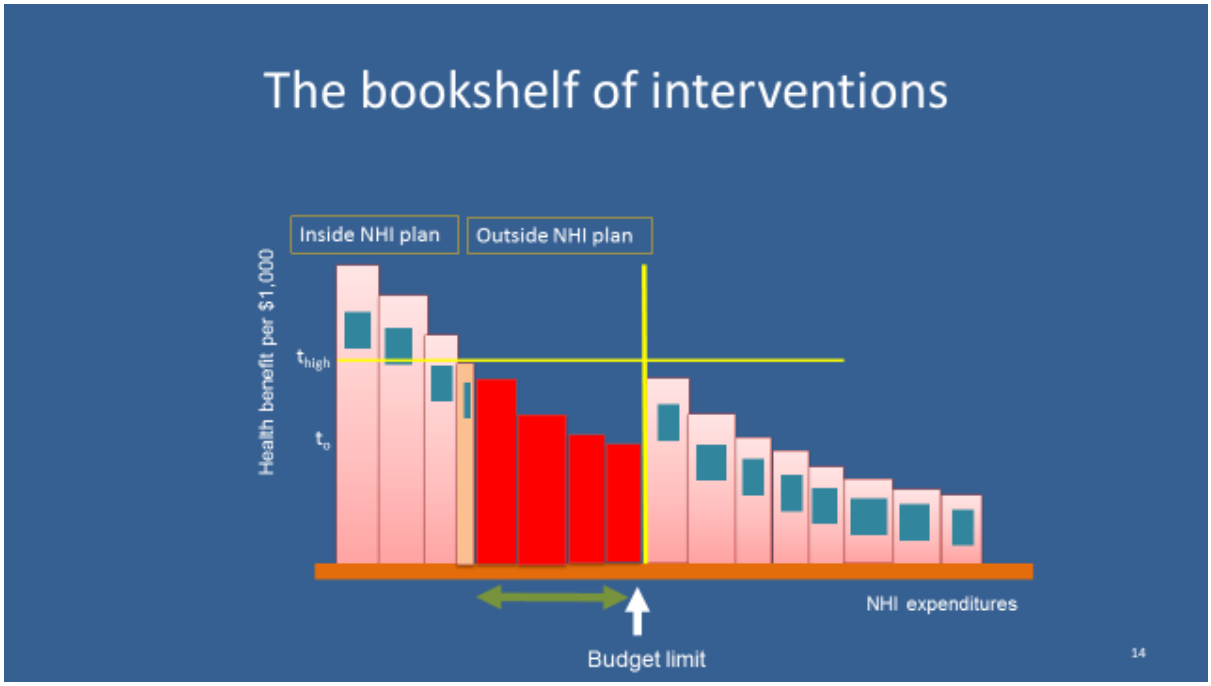


Figure 8 Health loss from threshold too high

that have yet to be approved, which are in direct competition with others yet to be approved, where all the gains are as yet only prospective and not as yet vested. It is nonetheless critically important to remember 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 publicly financed health care. Any other threshold is arbitrary and harmful to health, notwithstanding the interests (doubtless very vocal) that may be vested in it.

The 'true' 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! NICE has done it. The WHO has done it. I myself have done it. These thresholds all implicitly assume that the fraction of the national cake that should go on health and health care is much larger than it actually is. They are aspirational. 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 you thereby also set the threshold, or in setting the threshold you imply the budget. You ought never to set the one without realizing that in so doing you imply the other.

Multiple thresholds?

Some jurisdictions may in the short term have more than one threshold. For example, South Africa has in essence two parallel systems. The private insurance/private provision sector and the public insurance/public provision one. About 80% of expenditure goes on 20% of the population and 20% goes on 80% of the population. 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. 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 is lower than for 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

In England and Wales an interesting issue has arisen. 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. What has not been appreciated is that some of the losses imposed on the "anonymous" losers to whom I referred earlier, will be losses also for people at the end of life. The point is that, whatever the merits or otherwise of favouring such groups in considering the introduction of new interventions, similar groups exist elsewhere and are actually receiving current treatments for other conditions, and the consequences for them as they are denied services will need normally to be taken into account on grounds of consistency (Paulden et al. 2014, Claxton et a. 2015, McCabe et al. 2015).

What is *my* threshold?

Failure to set a threshold can have similar consequences to setting one that systematically admits too many low productivity interventions. If you deliberately fail to set a threshold you should probably stop pretending that you are trying to have the maximum impact on your people's health. Many countries are shy about being explicit about thresholds (Canada is one, so is 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. Canada is certainly well-

endowed in this respect. 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 the true threshold. Yet another approach, appropriate for countries with very low public expenditures, might be simply 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 rather 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. at York (Claxton et al. 2015). This econometric work does not make my simplifying assumption that no technology in use has a lower productivity than the threshold, and exploits the existence of programme budgets in the NHS. These cover 23 budgets. Changes in them can be linked to changes 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. Probabilistic sensitivity analysis indicated that there was a 0.89 probability that the true value lay below £20,000 per QALY and a 0.97 probability that it was less than £30,000 per QALY. This method tends to produce an underestimate of the 'true' threshold, as noted above.

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: judging what to include in the insured bundle cannot be solely based on evidence, even if it is good evidence (Culyer 2014). 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…

So, to return to my central theme of multum in parvo – much from little – we start with the simple idea that more health is a good thing; show that to achieve more health it must be possible to compare interventions in terms of their impact on a common measure of health; show that mere effectiveness is not a persuasive case for inclusion in public insurance plans, and that public health advocates need to address issues of relative effectiveness if they are to

be more effective advocates for public health. I have shown that a benchmark or threshold ratio of health gain to expenditure identifies the least effective intervention that should be included in a public insurance plan; that the reciprocal of this ratio – the cost-effectiveness threshold – should rise or fall as the health budget rises or falls (ceteris paribus); that setting thresholds too high or too low costs lives; that failure to set a cost-effectiveness threshold also involves avertable deaths and morbidity; that the threshold cannot be set independently of the health budget; that the threshold can be approached from either the demand-side or the supply side, the two being equivalent only in a health-maximising equilibrium; that the supply-side approach tends to underestimate the true threshold; that multiple thresholds are implied by systems having distinct and separable health budgets; that disinvestment involves eliminating *effective* technologies from the insured bundle; that anonymity and identity may be factors that affect the interpretation of the threshold; and (implicitly but all the way along) that the true opportunity cost of health care in not to be measured in money but in health itself.

That seems to me to be quite a lot to get from such a modest starting point. I hope Alchian and Williams might have approved.

References

Alchian, A. and W.R Allen (1963), *University Economics.* Belmont California, Wadsworth.

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.

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., C. McCabe, A. H. Briggs, K. Claxton, M. Buxton, R. L. Akehurst, M. Sculpher, J. Brazier. "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*, 2007, 12: 56-58.

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.

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

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

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.

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).

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

Williams. A. (1963), *Public finance and budgetary policy*, London, Allen & Unwin.

Williams, A. (1985), “The economics of coronary artery bypass grafting”, *British Medical Journal*, 291, 326–9.

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.

Acknowledgments

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, Mark Sculpher, Mike Paulden, Paul Revill, Adrian Towse and Yot Teerawattananon.