Understanding the Cost Effectiveness Threshold: a CASMI working paper

A working paper on the cost effectiveness threshold: understanding its function in Health Technology Appraisal, how it is estimated, and an international comparison of its value.

Initial paper: June 2016
Final paper: December 2017

Reference

Please cite this working paper as:

Morrell, Liz (2017). Understanding the cost effectiveness threshold: a CASMI working paper. CASMI, University of Oxford, UK. www.casmi.org.uk/

UNDERSTANDING THE COST-EFFECTIVENESS THRESHOLD

EXECUTIVE SUMMARY

The cost-effectiveness threshold is a central concept in the process of making decisions about whether a particular technology will be funded by the health service in the UK. It is therefore essential for anyone involved in drug access policy to understand the nature of the threshold, and how it relates to the country's health economy (for example: the health budget, inflation). In developing a policy position on changes to the threshold, it is also helpful to understand how its value is estimated, and how it compares with thresholds used in other countries.

This working paper was developed by CASMI during summer 2016, as part of a project exploring priorities for reform of NICE. It aimed to provide an overview of current thinking and research findings about the threshold as applied in decision-making in the UK, and addressed three topics:

Understanding the threshold

Section 1 presents an illustration of what the threshold represents, and how it varies with changes in the local health economy. The model illustrates that the threshold is not something that we *define*, but need to *estimate*. It is, theoretically, tightly linked to budget, inflation, productivity, and is dependent on what we measure as health gain. Although the direction of movement of the threshold for some simple scenarios is predictable (such as budget growth), in reality the factors are varying simultaneously and continuously, with interacting and countervailing pressures. It is therefore difficult to be definitive on the direction and extent of a change to today's threshold.

Empirical estimates of the UK threshold

Section 2 describes results from empirical studies to estimate a UK threshold, with supporting data from other relevant (mostly European) countries. The only significant study to attempt to quantify the actual threshold in the English NHS, gave a central estimate of £13k, ie well below the current NICE range. Attempts to estimate the value of a QALY based on the public's Willingness to Pay (WTP) gave estimates broadly consistent with NICE's threshold, although highly sensitive to study design. Estimates using the Value of a Prevented Fatality approach tend to be higher, particularly for life saving interventions. UK threshold estimates by these methods are in line with estimates for other European countries, particularly when adjusted for per-capita GDP. Overall, these empirical measures do not provide a clear-cut case to adjust the current UK threshold in either direction. Of note, most of the data were generated pre-2010.

International comparison

Section 3 shows thresholds used in comparator countries. The UK threshold is broadly in line with the generally accepted implied thresholds from other countries, although specific decisions may appear to extend those ranges. We note that NICE is unique in having an explicit threshold, although it is defined as a range and set within a deliberative process where it is not the sole determinant of a funding decision. Our understanding of other countries' thresholds is largely implicit or inferred. Countries that state they have an adjustable threshold, have a higher upper limit than the UK's. There have been no recent, explicit changes to thresholds made public, in the UK or elsewhere.

Policy implications

Adopting a policy position calling for a change in the threshold – in either direction – is challenging. The threshold estimate from measuring actual displacement in the NHS is unpalatably low, and although unlikely to be implemented, makes it difficult to argue for a threshold increase, particularly in the absence of a strong empirical case for a raise; further, the theoretical interpretation of the threshold does not provide unambiguous direction when multiple elements (budget, inflation, productivity) are simultaneously in flux. The fact that we have fairly insecure estimates as a start-point adds to the challenge of identifying a suitable basis to 'update' the existing value.

The lack of recent threshold updates – and the related data – reflects the challenges of estimating and implementing a change. Policy interest has shifted towards 'weighting' approaches, that allow decision-makers to adjust decisions that appear out of line with public wishes, such as at the end of life and for severe or rare conditions; this has also been the focus of more recent empirical research. Identifying such features that appear to attract public support, may be a more effective route to reform than calling for changes to the threshold itself.

Funding: the work was funded by the Policy Department, Cancer Research UK (CRUK). This working paper reflects evolving thinking during exploration of this topic in the summer of 2016. It is not necessarily the current policy position of either CRUK or CASMI.

UNDERSTANDING THE COST-EFFECTIVENESS THRESHOLD

BACKGROUND

The cost-effectiveness threshold is a central concept in the process of making decisions about whether a particular technology will be funded by the health service in the UK. This decision is based on cost effectiveness: how much health is created for the population for each pound spent.

The costs and effects of a new treatment or test are compared the current care provided, by calculating the *incremental cost effectiveness ratio*: the additional cost for the new treatment, divided by the additional health benefit. Health is usually measured using Quality-Adjusted Life Years (QALYs): a measure that takes account of both the duration of life, and the quality in which that life is lived – hence the cost effectiveness ratio is often referred to as the cost-per-QALY. This is then compared to a *threshold value* to determine whether the new treatment is considered to be cost-effective – that is, a good use of NHS resources.

The threshold range used by NICE, in making funding decisions for NHS England, is £20 000 to £30 000 per QALY. Below £20 000/QALY, a new technology will be considered cost effective, and will typically be recommended for funding. Technologies above £30, 000/QALY are generally not considered to be a good use of NHS resources, and would not normally be funded.

NICE's End-of-Life criteria define circumstances where its appraisal committees can in effect recommend technologies up to £50,000 per QALY, by applying a weighting of up to 2.5-fold to the QALY gain from the technology.

As the threshold is so critical in the decision process, it is essential for anyone involved in drug access policy to understand the nature of the threshold, and how it relates to the country's health economy (for example: the health budget, inflation). In considering a policy position on changes to the threshold, it is also helpful to understand how its value is estimated, and how it compares with thresholds used in other countries.

This working paper was developed by CASMI during summer 2016, as part of an exploration of priorities for reform of NICE. It aimed to provide an overview of current thinking and research findings about the threshold as applied in decision-making in the UK, and addressed three topics:

1. Understanding the cost effectiveness threshold
2. Empirical estimates of the UK threshold
3. International comparison of threshold values

[Funding: the work was funded by the Policy Department, Cancer Research UK (CRUK). This working paper reflects evolving thinking during exploration of this topic in the summer of 2016. It is not necessarily the current policy position of either CRUK or CASMI.]

SECTION 1. Understanding the cost effectiveness threshold

This section is based on a presentation developed to explain the concept of the cost-effectiveness threshold, and how it is expected to change with market circumstances, to an audience of policy advisers (ie not health economists, but familiar with health technology appraisal and questions of market access). The presentation slides are reproduced here as figures, to illustrate the narrative.

What is the cost-effectiveness threshold?

The cost-effectiveness threshold is relevant in a health care system with a fixed budget – as in the UK. It is an estimate of health forgone when existing NHS activities are displaced to accommodate additional costs of introducing new technologies (1).

Our startpoint is the assumption that the aim of healthcare is to maximise health for the population it serves. If a new treatment is introduced at higher cost than the current one, that necessarily means that funds must be diverted from elsewhere in the health system. This results in foregone health for other patients. If the new technology exceeds the cost-per-QALY threshold, that indicates the health loss by diverting funding exceeds the health gained from the new technology, thus resulting in net health loss for the population as a whole; this would normally not be considered a good use of health care resources, and the technology would not be funded.

The role of the threshold in effect is to represent the other patients in the health service, some of whom must inevitably bear the opportunity cost of diverting resources to fund a new intervention. The patients who will benefit from a new treatment are in the room at the Appraisal Committee, and their advocates have an opportunity to provide input to the appraisal from the scoping stage. The unidentified other patients are not present, and the threshold is the mechanism used by NICE to represent their interests (2).

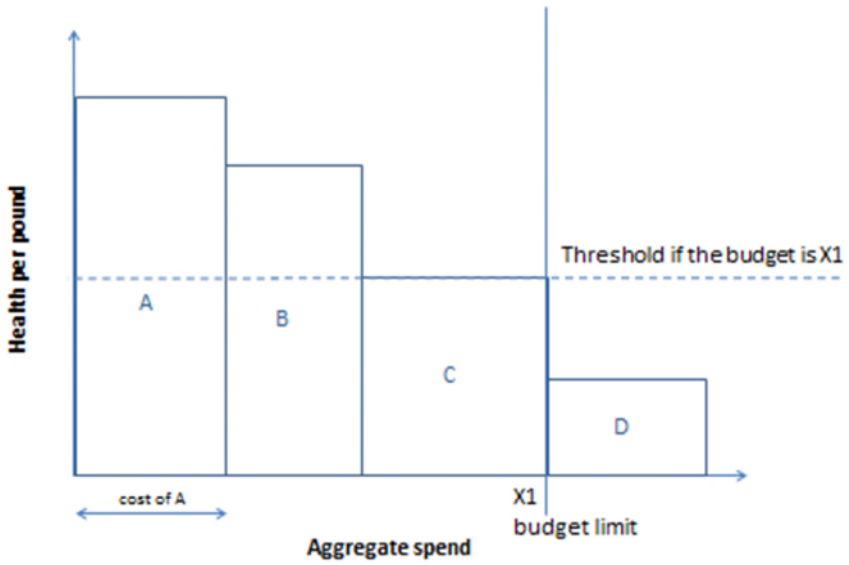
How could we derive the value of the threshold?

This illustration aims to provide a picture of what the threshold represents, and how it responds to changes in the health system. We assume that we can that we can measure health outcomes in a comparable way across therapeutic areas and interventions; this could be the QALY, or any an agreed measure of 'health'. The explanation below is adapted from (3). Rather than the cost per QALY, we're going to consider 'health gain per pound spent'; this is of course the cost effectiveness threshold turned upside down, and so it changes in the opposite direction (when health per pound goes up, the threshold goes down).

If we could, the ideal way to maximise health outcomes from a fixed NHS budget would be to rank all possible activities by health-per-pound, then implement them in order from the highest health-per-pound, until the budget X1 is used up, which in this illustration happens at project C (figure 1). Each activity is represented by a rectangle, whose *area* represents the amount of health gained; prioritising the activities in this way maximises the health (area of rectangles) from the budget.

If we then want to introduce a new intervention, it has to be at least as good as the last thing that's being funded (project C). If it's worse – such as project D - then it will generate less health (less area in the diagram) than whatever it displaces, and the health system will be getting less health out of its budget than with the original choices. Hence the health-per-pound of C has become the threshold. This is also known as the *shadow price of the budget constraint*.

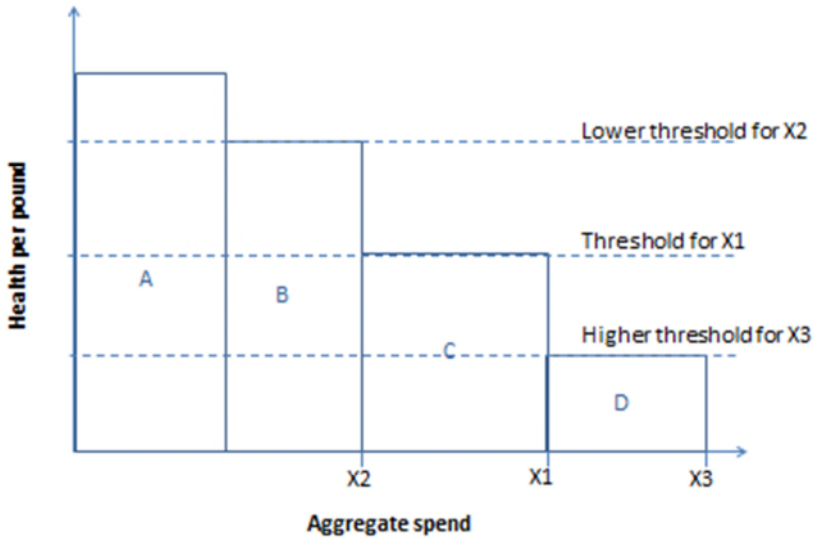
Figure 1. Identifying the threshold at budget of X1



We can use this figure to see how that threshold might change in various situations.

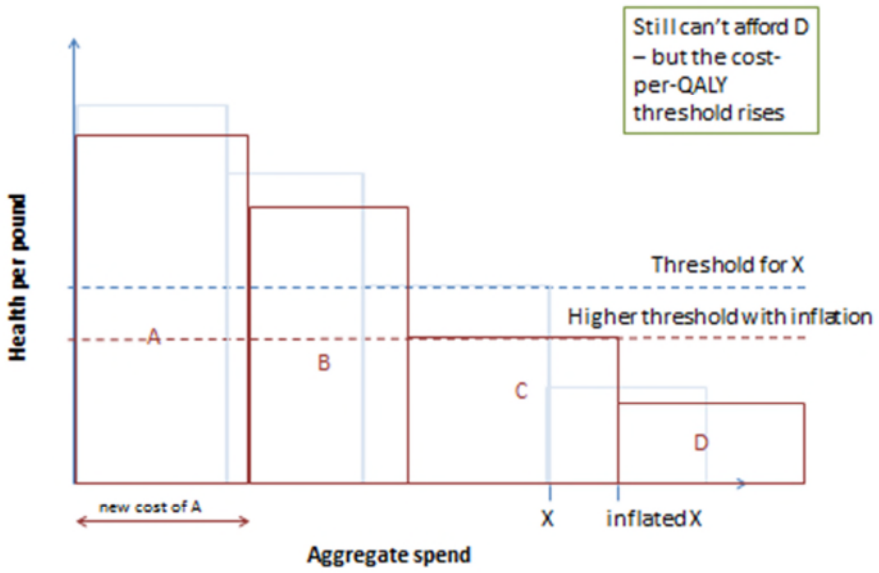
Budget increase (figure 2) – if the budget rises to X2, we can now fund project D, and the cut-off cost-per-pound falls (cost-per-QALY rises) to the value for D. Similarly if the budget falls to X3, we can only fund A-B, and the cut-off cost-per-pound rises (cost-per-QALY threshold falls) to the value for B. So in principle, the threshold follows the budget.

Figure 2. Changing the budget to X2 or X3



Inflation with corresponding budget growth (figure 3) – inflated values superimposed in red. Each activity cost more (wider rectangle), but the health gain (area) remains unchanged. We can still fund A-C, and the heath-per-pound falls (cost-per-QALY rises) to the new value for C post-inflation. But with budget growth in line with inflation, we still have no budget left over for D, which remains less cost effective than the interventions we're funding. So although in principle the threshold is more generous, in practice a new intervention like D may still not be able to displace C.

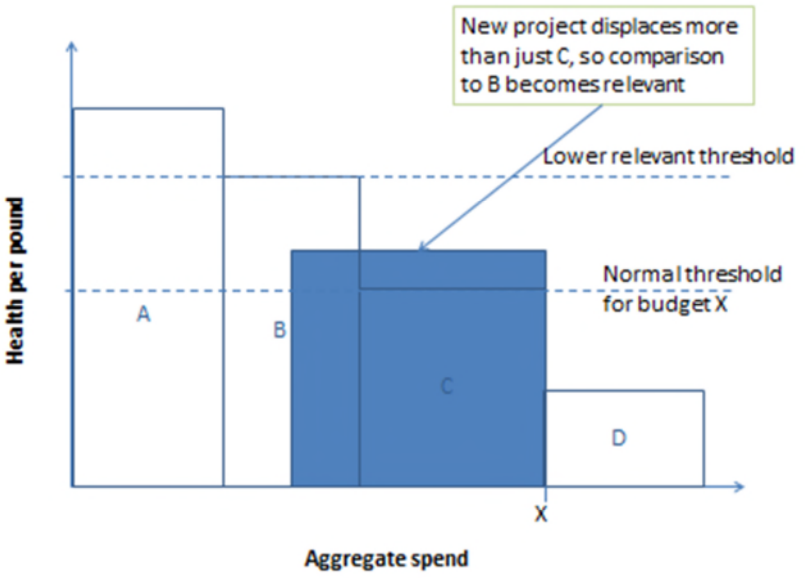
Figure 3. Effect of inflation with corresponding budget growth



A new intervention with major budget impact (figure 4) – the displacement effect of the blue shaded project extends beyond the 'last project' C, so we have to consider whether the new project is a better use of funds not only of C, but also B with its higher cost-per-pound (lower cost-per-QALY). Hence it has been argued that the threshold should be lower for such interventions; this could have applied, for example, to the hepatitis C medications.

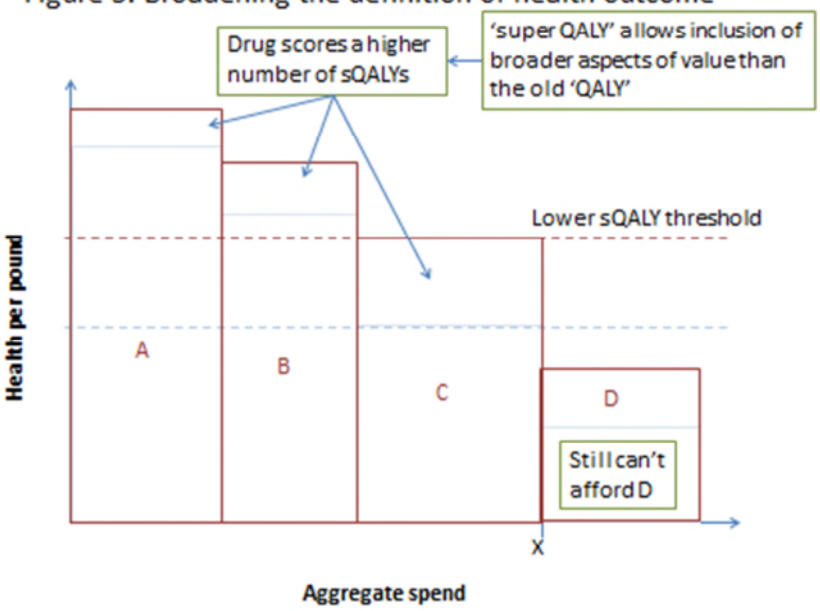
This is relevant in the context of the Budget Impact Test (BIT) introduced in April 2017. Under this test, for interventions with a budget impact of £20million in any of the first three years, NHS England can request an extension to the usual 90 day limit for implementing a NICE recommendation. The rationale is that this extension allows NHS England time to adjust budgets to incorporate the cost – ie it is entirely practical, addressing a completely different issue from that of the theoretically relevant threshold for high-budget-impact interventions.

Figure 4. Effect of an intervention with major budget impact



Broadening the definition of health outcome (figure 5) – we can consider broadening the elements included in the estimate of health outcome: for example to include productivity gains, impact on family and carers, or heath aspects not captured in EQ5D. Interventions are expected to accrue more of these 'superQALYs' (sQALYs) because of the broader definition, improving their apparent cost effectiveness. In figure 5 this is shown superimposed in red. However, it's not only wannabe intervention D that has more sQALYs - notice that rectangles A-C also have increased in height. As a result, C is still a better use of resources than D, and has defined a new higher cost-per-pound target (lower cost-per-sQALY threshold). The only way we can accept D is if it gains a lot of sQALYs, and C gains hardly any, so D overtakes C – that is, the sQALY captures elements that are specifically delivered by the new interventions.

Figure 5. Broadening the definition of health outcome



We could argue that Scotland's PACE meeting is doing just this – looking for unaccounted-for health gain for new drugs. Importantly, though, some unaccounted benefits will presumably accrue to current treatments, and if these are not considered, the process is implicitly overvaluing the new.

Productivity: cutting out completely ineffective activities – this is equivalent to a budget increase, if funds are freed up without a loss of health, so the cost-per-QALY threshold rises as in Figure 2.

Other scenarios are less clear to interpret theoretically. For example, **productivity improvements** (delivering a specific activity at lower cost with no change in outcome): as costs fall, health-per-pound rises (the rectangle keeps the same area, but become taller and thinner), so the cost-per-QALY threshold is expected to fall (1). However, its ultimate level depends on how the freed-up funds are spent, so potentially the threshold rises (4).

Similarly, if **inflation outstrips budget growth**, it can result in an effective budget decrease (cost-per-QALY threshold falls) or inflation of the cost of the 'last project' (cost-per-QALY threshold rises, but more budget consumed by current activities). The final position depends on how the system responds to the shortfall, and the relative impact of inflation on new technologies compared to existing care.

The further challenge is that all of these parameters are likely to be changing simultaneously, with countervailing pressures that are not readily predicted in this simple qualitative model.

Of course we could never do this exercise, nor do we disinvest rationally by QALY rules; disinvestment is decided, and is influenced, by local politics, equity considerations, and the need to hit mandatory targets (5). However it does illustrate the in-principle relationship between the threshold and the budget, inflation, productivity, and investment decisions - and why the role of NICE has been described as 'seeking' the threshold rather than setting it (6).

How can thresholds be derived?

Given we don't have the full comparable dataset to derive the threshold in this way, approaches to date have included:

- **League tables**: by comparing the ICER of a new intervention to programmes that are already funded, even if that comparison set is incomplete. This approach can be useful where a new intervention has an ICER lower than something accepted as cost-effective, or higher than something considered not cost-effective. An early example of this approach included the estimate of $50k/QALY renal dialysis believed to form the basis of Medicare's threshold. However the approach can create difficulties where the benchmarks considered "too high" are in fact in current use.

- **Implied threshold from previous decisions**: this approach has been used by decision-makers as well as researchers. It provides a useful benchmark, but does carry the assumption that previous decisions have been rational and objective, based on a similar budget constraint, and used comparable data and analyses; this in particular is likely to be violated in the absence of a Reference Case. The approach will also tend to reinforce the status quo.

- **Arbitrary "rules of thumb"**: these are often attractively round numbers but appear to have little empirical basis (2)

- **Multiplier of GDP**: the WHO CHOICE project suggested relating the cost of technologies to GDP, with interventions costing 1-3 times the country's per-capita GDP per Disability-Adjusted Life Year being considered cost-effective (7). This has the advantages of simplicity. However it does not take account of a society's willingness to pay for health care as a proportion of its GDP, and can lead to spiralling costs as it does not reflect opportunity cost. WHO have clarified that such a threshold is not intended to be used alone, but alongside other considerations such as affordability (7-10). Note that the UK's per capita GDP in 2016 was £33k ($41.6k, converted at 31/12/16 rate) – broadly in line with the current threshold.

Estimating the threshold is not a new problem; it was examined by the Commons Health Select Committee in 2008, and initial thinking on a method using NHS data was presented as part of the follow-up. Work by the Office of Health Economics reached expert consensus that there is no single agreed approach to estimating the threshold, but that we should use several complementary methods and take a triangulation approach to reach a reasonable estimate (11); most of the approaches suggested were based on measuring actual decisions in the NHS. The OHE work also acknowledges that the threshold is dynamic, changing as outlined above with budget, inflation, productivity, and investment decisions, so ideally would be estimated by methods allowing routine updating. A similar observation by Appleby *et al* (12) led to suggestion of a "threshold committee" to review and update the threshold.

A subtle distinction….

… between moving the threshold, and weighting the health gain (QALYs). This is particularly important when we consider differential thresholds – for example, having a higher threshold for serious disease, or for particular groups of patients.

Remember that the threshold represents the cost-per-QALY of what is displaced. If we raise the threshold for a particular group, then we are effectively down-weighting the QALY gain of the displaced treatment. In contrast if we weight the QALYs, we are explicitly giving a higher value to the QALY gain of the group under consideration. These are mathematically equivalent, but theoretically and philosophically quite different: if we adjust thresholds, the patients whose QALYs are being adjusted are not the beneficiaries of the drug under consideration, but a different group. Adjusting their QALYs, based on the characteristics of the beneficiaries, is a logical inconsistency (13). It also creates the difficulty that some patients in the displaced group may share the characteristic that is being given preference (for example, the opportunity cost for a drug approved under End-of-Life may be borne by other patients with short life expectancy elsewhere in the NHS) (14).

This is why NICE's End-of-Life adjustment carefully does not mention raising the threshold, but directs the Appraisal Committee to consider the weighting of QALYs that would be needed to make the intervention cost effective, and whether that value is plausible (15). Similar wording is applied in the weighting for the magnitude of health gain that was applied to the Highly Specialised Technologies appraisal process in April 2017 (16)

Policy Implications

The model illustrates that the threshold is not something that we define, but need to estimate. It is tightly linked to budget, inflation, productivity, and is dependent on what we measure as health gain. Although the direction of movement of the threshold for some simple scenarios is predictable (for example: budget growth, simple inflation), in reality the factors are varying simultaneously and continuously, with interacting and countervailing pressures. It is therefore difficult to be definitive on the direction and extent of any change to today's threshold. The fact that we have fairly insecure estimates as a start-point adds to the challenge of identifying a suitable basis to 'update' the existing value. There is currently no consensus on methods to estimate the threshold; recent thinking is that multiple approaches are needed to triangulate a feasible estimate.

SECTION 2: EMPIRICAL ESTIMATES OF THE UK THRESHOLD

Given the central role of the threshold in decision-making, it is perhaps surprising that NICE's current threshold was implied, from decisions made by NICE prior to 2004; it has not been changed since then, and is widely acknowledged to have limited empirical basis (2, 17). Estimating an empirical value for the threshold has proved challenging.

Table 1 summarises results of studies from the UK and other developed countries, identified from a review of the literature. There have been two main types of approach, as described below. The first approach is often referred to as a 'supply-side' approach, as it considers the 'supply' of healthcare, reflecting the ability of the healthcare system to generate health. Correspondingly, the second approach is 'demand-side', as it considers the value placed on healthcare by the population, and hence the demand for care.

The findings of this review are consistent with those reported in a published study (18) which was published since the original draft of our working paper; it is based on a broader literature review, and comes to very similar conclusions.

1. Estimating the effect of expenditure change. Work by Claxton *et al* at the University of York (1) aims to measure the actual threshold in the NHS – ie how spending changes in response to budgetary changes, and the resulting impact on health outcomes. It is the first attempt to measure the revealed threshold, and was funded by the National Institute for Health Research (NIHR).

Results indicate that the current NICE range of £20-30k is too high, with a central estimated value of **£13k** (an earlier version of the work suggested £18k, but this was re-estimated following refinements to the model). The work has been criticised for the number and scale of assumptions required (4). However it appears unlikely that this model, even with refinement of the assumptions and inputs, will support recommendations to increase NICE's threshold dramatically above the current range. Given the debate about the certainty of this estimate, and concerns over the political feasibility of implementing it, there does not appear to be appetite at this time to adjust the threshold downwards in response (17, 19).

[Update January 2018: a recently published paper attempted a similar estimate for Australia, and also found a displacement cost-per-QALY well below the threshold in use for decision-making (20) (AU$28k, compared to an implicit threshold of ~AU$50k).]

The study in Scotland by Karlsberg Schaffer et al (5) is based on actual budget decisions made by NHS Boards, assuming that the services seeing significant changes in funding are the 'marginal' services that represent the threshold. These services were identified from public domain audit reports and the cost-per-QALY estimated from literature searches. These estimates varied widely, so it was not possible to provide a reliable value for the threshold.

2. Estimating the value of a QALY. The European 'Social Value of a QALY' (EuroVaQ) project, which involved researchers from multiple universities, tested two parallel approaches: estimates based on the value of preventing a transport fatality (VPF) (21), and measures of public willingness-to-pay (WTP) to improve specified health states (22). Both of these use 'stated preference' methods; they ask respondents hypothetically how they value

something, in contrast to observing their actual behaviour, which would be a 'revealed preference' approach.

The *fatality prevention* approach gave different values depending on the type of health gain, with life-saving being worth more than life-extending or improvements in quality of life. The result for life-extending treatments is £35k/QALY, similar to the NICE range but below the implied range for end-of-life treatments. Of note, the results for life-saving interventions cannot be applied directly to cancer patients with limited life expectancy, as the study scenario assumed that the 'saved' patients would go on to live healthy lives with a normal life expectancy, which is not typically the case for cancer treatments (21).

Results in the *willingness-to-pay* studies were highly sensitive to research design, but gave results for the UK broadly consistent with the current threshold range, although extending below the lower limit. WTP estimates are typically lower than the VPF values (18).The UK values were in line with those from France and Netherlands, and a little below Scandinavian countries and Spain; however when the values were adjusted for per-capita GDP, Spain was a high outlier along with Poland and Hungary, with the UK in line with the other countries (data not reproduced here - see (23)).

Note that social values derived in this way are considered to be more relevant to the setting of the health care budget than to defining the threshold directly, as they are based on trade-offs people make between health care, other public sector expenditure, and personal consumption (ie income after tax) rather than opportunity cost in terms of health (6, 24). By this approach, logic demands that all interventions falling below the societal WTP should be adopted, and the budget adjusted accordingly. In contrast in health systems like the UK with a fixed budget, displacement is inevitable, and the threshold approach is more appropriate. If the actual threshold is dramatically lower then societal WTP, it can be argued that the health budget is set too low.

Limitations: All of these studies are based on data generated between ~2005 and 2010. This reflects a shift in research interest towards defining weights for specific types of patient or health gain, for example as stimulated by the Value-Based Pricing work in the UK in 2010-2014. It could therefore be argued that the data are in need of updating, or at least adjusting to today's prices.

Further, given the sensitivity to study design, it is difficult to regard these numbers as an absolute value, and they are perhaps better used comparatively to look at relative values between patient groups.

Policy implications

Arguing for a change in threshold in either direction is likely to meet with resistance from some stakeholders; the York result is unpalatably low, and although unlikely to be implemented, does make it difficult to argue for a threshold increase, particularly in the absence of a strong empirical case.

There is a case for further work to update and extend the previous studies; in particular, for taking different approaches that consider the threshold from different perspectives, with the aim that the combined evidence will indicate future direction for threshold evolution.

Table 2. Empirical estimates of the cost effectiveness threshold

COUNTRY	METHOD	SUPPLY- or DEMAND?	YEAR OF DATA/ FIRST PUBLICATION	VALUE PER QALY
UK	NHS England: budget and outcome elasticity (1)	supply	2008-2010/2013	£13k (£17k in cancer)
	NHS Scotland: marginal disinvestment decisions (5)	supply	2012/2013	Wide variation
	Value of Prevented Fatality/Injury (21)	demand	2005/2009	£70k (life saving) £35k (life extending) £10k (QoL enhancing)
	Mean WTP with serious illness to gain 1 year in good health (25)	demand	2007/2010	£23-38k
	Mean WTP for vignettes valued by standard gamble (22)	demand	2010/2013	$13-29k* (£9-20k)
Netherlands	Mean WTP with Dutch EQ5D tariff, improving QoL (26)	demand	2008/2010	€24.5k
	Mean WTP with EQ5D VAS, improving QoL (26)	demand	2008/2010	€12.9k
	Mean WTP societal or socially inclusive perspective, improving QoL (27)	demand	2010/2013	€52-83k
	Mean WTP for vignettes valued by standard gamble (22)	demand	2010/2013	$16-27k*
Sweden	Value of avoiding a transport fatality (28)	demand	2003 (value in 2006 prices)	€90k
	Pilot study WTP for insurance (28)	demand	?	€40k
	Mean WTP for vignettes valued by standard gamble (22)	demand	2010/2013	$18-35k*
France	Mean WTP for vignettes valued by standard gamble (22)	demand	2010/2013	$11-27k*
Spain	Mean WTP for vignettes valued by standard gamble (22)	demand	2010/2013	$26-53k*
Australia	Mean WTP with serious illness to gain 1 year in good health (25)	demand	2007/2010	AU$64-89k
	Australian healthcare: budget and outcome elasticity (20)	supply	2012/2017	AU$28k (£16k[#])

Table 2:

WTP: willingness to pay VAS: visual analogue scale
* converted to US$ from local currency studies, at 2008 Purchasing Power Parity conversion rates
# converted at January 2018 exchange rate £ = AU$1.73

SECTION 3. INTERNATIONAL COMPARISON OF THRESHOLDS

To place the UK's threshold in context, we compared the UK's philosophy on the threshold and its value with those in use in other countries. Information was obtained from a review of published literature, government documents, and discussions with experts involved in Health Technology Appraisal; findings are summarised in Table 3 and discussed below.

Overall, the data suggest that UK threshold values are slightly, but not dramatically, below the other countries listed when converted to a common currency. However, note that the values are not adjusted for Purchasing Power Parity, so may not fully reflect how the local population view that number relative to their own wealth. Note also that the currency conversions were originally done before the Brexit referendum; the Sterling equivalents for Euro and dollar values would be higher at 2017 exchange rates.

NICE is unique in having an explicit threshold range, although it is framed in the context of a deliberative process in which, when ICERs are above this range, "the Committee will need to identify an increasingly stronger case for supporting the technology as an effective use of NHS resources" (13). The threshold is explicitly described as *not* being the sole determinant of a decision. AWMSG in Wales uses very similar wording at the same threshold limits (26). Many other countries do not specify a threshold, and the generally accepted understanding of those thresholds is inferred from past decisions, comments from key players, or suggestions in government documents.

Because NICE has an explicit threshold, this is the number that is most discussed. For comparability with the other countries, we should also consider an inferred threshold for NICE. A recent study indicated that NICE recommends technologies above 30k, with a 50% probability of acceptance at £39-44k/QALY (excluding end-of-life drugs) (29), broadly consistent with those seen in other countries. CASMI's work on uncertainty also observed several examples where NICE recommended drugs with ICERs above £30k, or indeed where no ICER could be estimated (30). Taken together, these observations indicate the use of other elements in the decision process, rather than absolute reliance on the threshold range.

Inference from past decisions is limited by the prevalence of confidential commercial arrangements, which obscure the actual price paid and result in an inflated estimate of an acceptance threshold. It will also tend to embed existing beliefs. Estimates based on the highest known examples are also likely to be inflated, as they may not represent typical decisions, and may have specific reasons for funding that are not generalizable.

The upper limits of the ranges are somewhat higher in Sweden and the Netherlands, both which claim to have 'moving' thresholds. Sweden specifically states that TLV's acceptance varies with severity of disease, which is operationalised by descriptive classification (low to high severity) (31). Decision data show that drugs for severe conditions were accepted at ICERs at which drugs for less severe conditions were rejected (28). The Netherlands have explored adjusting their threshold according to the proportional shortfall of QALYs caused by the disease (31, 32); however we have to date found no published evidence of how this has worked in practice nor its impact on access.

The origin of the Medicare dialysis $50,000/QALY is difficult to trace through the literature. Following the decision to mandate coverage of renal dialysis in the 1970s, the figure makes an appearance in the early work using league tables in the 1980s, and gained popularity in analyses in the 1990s, although more recent analyses also refer to benchmarks of $100,000. The history suggests it represents a cost-per-QALY implied to be acceptable, rather than a cut-off above which is unacceptable.

One observation from the table is that the assumed thresholds, whether explicit or implicit, have not changed over time - at least not as documented in the public domain. We might therefore expect arguments from some stakeholders that a revision is appropriate.

Policy implications

The thresholds in routine use appear generally comparable in the countries considered. Differences in funding decisions may be driven not so much by the threshold itself, but by countries' ability to be flexible in specific cases, such as drugs for severe or rare conditions, or use at the end of life. Identifying such features that appear to attract public support, may be a more effective route to reform than calling for changes to the threshold overall.

Table 3. International comparison of the use of thresholds

COUNTRY	BASIS	YEAR OF ESTIMATE	THRESHOLD	STERLING EQUIVALENT*
England	• £20-30 000 per QALY. Deliberative process with increasingly stronger case required with higher ICERs (15) • End-of-Life criteria allow consideration of weighting required to bring the ICER into cost-effectiveness range; generally accepted maximum of ~2.5 fold • Inferred threshold £39-44k (50% probability of acceptance) (29) • Highly Specialised Technologies: £100 000 per QALY, with up to three-fold weighting for large QALY gains (16)	2008, 2013 2009 2015 2017	£20-30k (£50k EoL) (£100K HST)	£20-30k (£50k EoL) (£100K HST)
Scotland	• "SMC does not have a formal threshold cost per QALY" (33). Analysis of past decisions suggests threshold similar to NICE. • End-of-life and orphan/ultra-orphan criteria implicitly allow for higher ICER based on additional information from PACE meeting; CASMI data show acceptance to £60k	2015 2015	£20-30k (£60k EoL/rare)	£20-30k (£60k EoL/rare)
Wales	• "AWMSG / NMG do not use a fixed ICER threshold" but describes the same deliberative process as NICE at £20-30k (34) • End-of-Life as NICE, orphan/ultra-orphan as Scotland.	2015	£20-30k	£20-30k (£50k EoL)
Australia	• Not explicit but assumed to be ~AU$50k (35) • Inferred AU$52200 above which reimbursement unlikely (36) • PBAC quote: AU$50k "on the high side" (37)	2015 2005 2009	AU$50k	£25k
New Zealand	• PHARMAC has no threshold, but considers proposals relative to budget (38) • Express as QALYs gained per NZ$million (35)	2015	-	-
Canada	• "the framework reinforces that there is no threshold that must be met for any single element in the review" (39) • CAD$50k was proposed in 1992; recent approvals up to CAD$144k in oncology (40)	2015 1992, 2015	CAN$50k, CAN$144k	£24k, £70k

Table 3 (continued) International comparison of the use of thresholds

COUNTRY	BASIS	YEAR OF ESTIMATE	THRESHOLD	STERLING EQUIVALENT*
Ireland	• Framework Agreement between the Department of Health and the Irish industry association in 2012 had an explicit threshold of €45K, replacing an unofficial threshold of €20K. Revised Framework Agreement in 2016 combines threshold range €20-45k with budget impact to determine authority level needed for the decision.(41)	2012	€20-45k	£15-33k
Netherlands	• Suggested range of €20-80k in ministerial documents but not publicly confirmed or endorsed (32)	2006, 2013	€20-80k	£15-59k
Sweden	• Highest approved is €90k for high severity; rejections at lower ICERs eg €40k for low severity. Benchmarks include NICE, and a VPF estimate (28)	2006	€40-90k	£29-66k
Belgium	• Inferred threshold from past decisions of €80k although some exceptions (42) • No formal threshold, but interviews indicated it would increase with severity and medical need (43)	2013	€80k	£59k
Austria	• No formal threshold, but interviews indicated it would increase with severity and medical need (43)			

* converted at Jan 1 2016 pre-Brexit rates. €=1.36 AU$=2.02 CAN$=2.05 U$=1.48
PACE: Patient and Clinician Engagement VPF: Value of a Prevented Fatality

REFERENCES

1. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess. 2015; 19.
2. Claxton K, Sculpher M, Palmer S, et al. CAUSES FOR CONCERN: IS NICE FAILING TO UPHOLD ITS RESPONSIBILITIES TO ALL NHS PATIENTS? Health Economics. 2015; 24: 1-7.
3. Drummond M, Sculpher M, Torrance G, et al. Methods for the Economic Evaluation of Health Care Programmes. 3rd ed. New York: Oxford University Press, 2005.
4. Barnsley P, Towse A, Karlsberg Schaffer S, et al. Critique of CHE Research Paper 81: Methods for the Estimation of the NICE Cost Effectiveness Threshold. In: Office of Health Economics, ed., Occasional Papers. https://www.ohe.org/publications/critique-che-research-paper-81-methods-estimation-nice-cost-effectiveness-threshold, 2013.
5. Karlsberg Schaffer S, Sussex J, Devlin N, et al. Local health care expenditure plans and their opportunity costs. Health Policy. 2015; 119: 1237-44.
6. Culyer A, McCabe C, Briggs A, et al. Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. Journal of Health Services Research & Policy. 2007; 12: 56-58.
7. Marseille E, Larson B, Kazi D, et al. Thresholds for the cost–effectiveness of interventions: alternative approaches. Bulletin of the World Health Organization. 2015; 93: 118-24.
8. CHE. iDSI Workshop on Cost-Effectiveness Thresholds: Conceptualisation and Estimation. In: University of York, ed. https://www.york.ac.uk/media/che/images/news/iDSI_Summary_Report-FINAL.pdf, 2015.
9. Garner S, Rintoul A, Hill SR. Value-Based Pricing: L'Enfant Terrible? PharmacoEconomics. 2017.
10. Bertram MY, Lauer JA, De Joncheere K, et al. Cost–effectiveness thresholds: pros and cons. Bulletin of the World Health Organization. 2016; 94: 925-30.
11. Karlsberg Schaffer S, Cubi-Molla P, Devlin N, et al. Shaping the Research Agenda to Estimate Relevant Cost-effectiveness Thresholds for Health Technology Assessment Decision Making: Report for ABPI. In: Economics OoH, ed., Consulting Reports. https://www.ohe.org/publications/shaping-research-agenda-estimate-cost-effectiveness-thresholds-decision-making, 2016.
12. Appleby J, Devlin N, Parkin D. NICE's cost effectiveness threshold. BMJ : British Medical Journal. 2007; 335: 358-59.
13. Wailoo A, Tsuchiya A, McCabe C. Weighting must wait: incorporating equity concerns into cost-effectiveness analysis may take longer than expected. Pharmacoeconomics. 2009; 27: 983-9.
14. Soares MO. Is the QALY blind, deaf and dumb to equity? NICE's considerations over equity. British medical bulletin. 2012; 101: 17-31.
15. NICE. Guide to the Methods of Technology Appraisal. http://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf 2013.
16. NICE. Interim Process and Methods of the Highly Specialised Technologies Programme: Updated to reflect 2017 changes. https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-highly-specialised-technologies-guidance/HST-interim-methods-process-guide-may-17.pdf, 2017.
17. Raftery JP. NICE's Cost-Effectiveness Range: Should it be Lowered? PharmacoEconomics. 2014; 32: 613-15.
18. Vallejo-Torres L, García-Lorenzo B, Castilla I, et al. On the Estimation of the Cost-Effectiveness Threshold: Why, What, How? Value in Health. 2016; 19: 558-66.
19. Dillon A. Carrying NICE over the threshold. www.nice.org.uk/news/blog/carrying-nice-over-the-threshold, 2015.
20. Edney LC, Haji Ali Afzali H, Cheng TC, et al. Estimating the Reference Incremental Cost-Effectiveness Ratio for the Australian Health System. PharmacoEconomics. 2017.

21. Donaldson C, Baker R, Mason H, et al. The social value of a QALY: raising the bar or barring the raise? BMC Health Services Research. 2011; 11: 1-8.
22. Robinson A, Gyrd-Hansen D, Bacon P, et al. Estimating a WTP-based value of a QALY: The 'chained' approach. Social Science & Medicine. 2013; 92: 92-104.
23. Nimdet K, Chaiyakunapruk N, Vichansavakul K, et al. A Systematic Review of Studies Eliciting Willingness-to-Pay per Quality-Adjusted Life Year: Does It Justify CE Threshold? PLoS ONE. 2015.
24. McCabe C, Claxton K, Culyer AJ. The NICE Cost-Effectiveness Threshold: What it is and What that Means. PharmacoEconomics. 2008; 26: 733.
25. Shiroiwa T, Sung YK, Fukuda T, et al. International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? Health Economics. 2010; 19: 422-37.
26. Bobinac A, van Exel NJA, Rutten FFH, et al. Willingness to Pay for a Quality-Adjusted Life-Year: The Individual Perspective. Value in Health. 2010; 13: 1046-55.
27. Bobinac A, van Exel N, Job A, et al. VALUING QALY GAINS BY APPLYING A SOCIETAL PERSPECTIVE. Health Economics. 2013; 22: 1272-81.
28. Persson U. Value Based Pricing in Sweden: lessons for design? In: Office for Health Economics, ed., Seminar Briefing. https://www.ohe.org/publications/value-based-pricing-sweden-lessons-design: Office for Health Economics, 2012.
29. Dakin H, Devlin N, Feng Y, et al. The Influence of Cost-Effectiveness and Other Factors on Nice Decisions. Health Economics. 2015; 24: 1256-71.
30. Morrell L, Wordsworth S, Schuh A, et al. Will the reformed Cancer Drugs Fund address the most common types of uncertainty? An analysis of NICE cancer drug appraisals. (submitted). 2017.
31. Franken M, Stolk E, Scharringhausen T, et al. A comparative study of the role of disease severity in drug reimbursement decision making in four European countries. Health Policy. 2015; 119: 195-202.
32. van de Wetering EJ, Stolk EA, van Exel NJA, et al. Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall. The European Journal of Health Economics. 2013; 14: 107-15.
33. Scottish Medicines Consortium. A Guide to the Scottish Medicines Consortium. https://www.scottishmedicines.org.uk/About_SMC/What_we_do/SMC_Guide__web___final_.pdf 2015.
34. AWMSG. AWMSG Summary Guidelines for Appraising Medicines. http://www.awmsg.org/docs/awmsg/appraisaldocs/inforandforms/AWMSG%20summary%20guidelines%20for%20appraising%20medicines.pdf, 2012.
35. Lorgelly P. From the Antipodes to the Motherland: Reflections on HTA Decision Makers as Budget Takers and Budget Makers. In: Office of Health Economics, ed., Lunchtime Seminars. https://www.ohe.org/news/antipodes-motherland-reflections-hta-decision-makers-budget-takers-and-budget-makers, 2015.
36. Henry D, Hill S, Harris A. Drug Prices and Value for Money: The Australian Pharmaceutical Benefits Scheme. JAMA. 2005; 294: 2630-32.
37. Lopert R. Evidence-Based Decision-Making Within Australia's Pharmaceutical Benefits Scheme. http://www.commonwealthfund.org/~/media/files/publications/issue-brief/2009/jul/chalkidou/1297_lopert_cer_australia_issue_brief_724.pdf: The Commonwealth Fund, 2009.
38. PHARMAC. Guidelines for Funding Applications to PHARMAC. https://www.pharmac.govt.nz/assets/guidelines-for-funding-applications-2015-08.pdf, 2015.
39. CADTH; pCODR. pCODR Expert Review Committee Deliberative Framework. https://www.cadth.ca/sites/default/files/pcodr/The%20pCODR%20Expert%20Review%20Committee%20(pERC)/pcodr_perc_deliberative_frame.pdf, 2016.

40. Husereau D, Culyer AJ, Neumann P, et al. How do Economic Evaluations Inform Health Policy Decisions for Treatment and Prevention in Canada and the United States? Applied Health Economics and Health Policy. 2015; 13: 273-79.
41. O'Mahony JF, Coughlan D. The Irish Cost-Effectiveness Threshold: Does it Support Rational Rationing or Might it Lead to Unintended Harm to Ireland's Health System? PharmacoEconomics. 2016; 34: 5-11.
42. Paris V, Belloni A. Value in Pharmaceutical Pricing. OECD Health Working Papers,. http://www.oecd-ilibrary.org/docserver/download/5k43jc9v6knx.pdf?expires=1467998407&id=id&accname=guest&checksum=ACE1856721CF50E9308D049E3DA692E8, 2013.
43. Franken M, le Polain M, Cleemput I, et al. SIMILARITIES AND DIFFERENCES BETWEEN FIVE EUROPEAN DRUG REIMBURSEMENT SYSTEMS. International Journal of Technology Assessment in Health Care. 2012; 28: 349-57.