

How to design the cost-effectiveness appraisal process of new healthcare technologies to maximise population health: A conceptual framework

Kasper M Johannesen, Karl Claxton, Mark J. Sculpher and Allan J Wailoo

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-140899>

N.B.: When citing this work, cite the original publication.

Johannesen, K. M, Claxton, K., Sculpher, M. J., Wailoo, A. J, (2017), How to design the cost-effectiveness appraisal process of new healthcare technologies to maximise population health: A conceptual framework, *Health Economics*. <https://doi.org/10.1002/hec.3561>

Original publication available at:

<https://doi.org/10.1002/hec.3561>

Copyright: Wiley (12 months)

<http://eu.wiley.com/WileyCDA/>



How to design the cost-effectiveness appraisal process

How to design the cost-effectiveness appraisal process of new health care technologies to maximise population health: A conceptual framework

Kasper M Johannesen¹, Karl Claxton², Mark J Sculpher² and Allan J Wailoo³

¹Center for Medical Technology Assessment, Linköping University, Sweden

²Centre for Health Economics, University of York, UK

³Health Economics and Decision Science, University of Sheffield, UK

Keywords: Economic Evaluation; Health Technology Appraisal; Cost-effectiveness appraisal and Decision Making.

How to design the cost-effectiveness appraisal process

Abstract

This paper presents a conceptual framework to analyse the design of the cost-effectiveness appraisal process of new health care technologies. The framework characterises the appraisal processes as a diagnostic test aimed at identifying cost-effective (true positives) and non-cost-effective (true negative) technologies. Using the framework, factors that influence the value of operating an appraisal process, in terms of net gain to population health, are identified. The framework is used to gain insight into current policy questions including: i) how rigorous the process should be; ii) who should have the burden of proof; and iii) how optimal design changes when allowing for appeals, price reductions, resubmissions and re-evaluations.

The paper demonstrates that there is no *one* optimal appraisal process and the process should be adapted over time and to the specific technology under assessment. Optimal design depends on country-specific features of (future) technologies, e.g. effect, price and size of the patient population, which might explain the difference in appraisal processes across countries. It is shown that burden of proof should be placed on the producers and that the impact of price reductions and patient access schemes on the producer's price setting should be considered when designing the appraisal process.

How to design the cost-effectiveness appraisal process

1. Introduction

Cost-effectiveness analysis has become a key decision criterion in reimbursement and implementation decisions for new health care technologies in many countries (Heintz et al., 2016; Kanavos et al., 2011). Determining the cost-effectiveness of new technologies is often the responsibility of governmental agencies like, for example, the National Institute for Health and Care Excellence (NICE) in England and Wales, The Scottish Medicines Consortium (SMC) in Scotland and The Dental and Pharmaceutical Benefits Agency (TLV) in Sweden. These agencies mainly rely on analyses performed and submitted by producers of a new technology, which are then reviewed by the agencies or third party evaluation groups (Sculpher, 2010; Kanavos et al., 2010). A key reason for agencies not relying solely on producer submissions is that producers have incentives to overestimate the cost-effectiveness of new technologies to improve the likelihood of achieving reimbursement, and producer cost-effectiveness estimates has been shown to differ significantly from independent assessment estimates (Barbieri et al., 2009; Miners et al., 2005; Bell et al., 2006; Chauhan et al., 2007).

Agencies have limited time and resources available to conduct their appraisals but this varies across countries. The SMC has 18 or 22 weeks and TLV has 180 days to conduct their own internal review and to make a reimbursement decision; whereas, under current arrangements, NICE has 35 weeks, under the Single Technology Appraisal (STA) process, to publish a recommendation, which includes the external “Evidence Review Group” (ERG) assessment¹ (NICE, 2014; TLV, 2016). Restricting the time and resources that such agencies have available to conduct their reviews is likely to reduce the precision of the cost-effectiveness assessment. Hence, there is a trade-off between the time and resources available to appraisal organisations

¹ The NICE STA timelines are changing for technologies submitted for review after April 1st 2017 undergoing the shorter “fast track” process or “commercial discussions” between the NHS England and the company (NICE, 2017a; NICE, 2017b).

How to design the cost-effectiveness appraisal process

and the accuracy of the cost-effectiveness appraisal process, in terms of correctly identifying cost-effective and non-cost-effective technologies (Bell, 2015; Kanavos et al., 2010; Kaltenthaler et al., 2008; Barham, 2008). There is limited research on this trade-off between resources expended on and accuracy of cost-effectiveness appraisals.

The purpose of this paper is to set up an analytic framework that enables analysis of different ways to design the cost-effectiveness appraisal of new technologies. The framework will be used to analyse and address some policy-relevant considerations including: i) How should the process be designed to maximize population health; and ii) what factors influence how to design the process optimally.

The next section outlines the framework where the cost-effectiveness appraisal is viewed as a diagnostic test aimed at identifying cost-effective technologies (true-positives) and non-cost-effective technologies (true-negatives). Section 3 utilises the framework to analyse how to design and operate an appraisal process optimally and to make some policy relevant conclusions. The paper ends with a discussion and concluding remarks.

2. The conceptual framework

Within healthcare, cost-effectiveness is often assessed by comparing the incremental cost-effectiveness ratio (ICER) with the cost-effectiveness threshold (λ), which represents the marginal productivity of the health care sector (Drummond et al., 2015; Claxton et al., 2010). More effective technologies with an ICER below λ , as in Equation 1 below (where ΔC and ΔE represents incremental cost and effect, respectively), will have a positive impact on population health if implemented and *vice versa* (Briggs et al., 2006; Claxton et al., 2015).

$$ICER < \lambda \Leftrightarrow \frac{\Delta C}{\Delta E} < \lambda \quad (1)$$

How to design the cost-effectiveness appraisal process

2.1. True- vs. claimed-ICER

Technologies with higher ICERs and ICERs above the λ have a higher likelihood of being deemed non-cost-effective and rejected² (George et al., 2001; Rawlins & Culyer, 2004; Devlin & Parkin, 2004; Dakin et al., 2015). Producers therefore have an incentive to claim that the ICER is below λ , even if this is not the case. Producers are able to claim lower ICERs by claiming higher ΔE and/or lower ΔC through optimistic assumptions, selective analysis of data or otherwise utilising their informational advantage about the new technology. Here the *claimed-ICER* is defined as the ICER claimed by producers in their reimbursement submissions; and the *true-ICER* is defined as the ICER based on true or unbiased estimate of ΔE and ΔC . Figure 1 presents possible combinations of claimed- and true-ICERs.

Figure 1 about here

The 45° line shows where the claimed-ICER is equal to the true-ICER. Truly cost-effective technologies will be located to the left of the vertical λ line (in areas 1, 2 or 3) and non-cost-effective technologies will be located to the right (in areas 4, 5 or 6). It seems unlikely that producers will, in general, put forward a claimed ICER above the horizontal λ line, i.e. indicate that a technology is non cost-effective. It is, therefore, to be expected that almost all technologies undergoing appraisal will have a claimed-ICER below λ and be located in area 2, 3 and 4.

Ideally decision makers would identify whether each technology lies to the left or the right of the vertical λ and base their decisions to reject or approve on this information. However, the true-ICER is never known with complete certainty at the time of cost-effectiveness appraisal, or at any other time, due to uncertainty around the costs and effects of implementing of a new technology into clinical practice. Also, appraisal organisations are generally unable to postpone

² We restrict our analysis to more effective technologies, i.e. $\Delta E > 0$, with no loss of generality.

How to design the cost-effectiveness appraisal process

reimbursement decisions or commission further research to get a better estimate of the true-ICER (Griffin et al., 2011; McCabe et al., 2010; McKenna et al., 2015). Appraisal organisations must, therefore, base their decisions on the submitted claimed-ICER and their own best estimate of the ICER, i.e. the *appraisal-ICER*, at the time of the appraisal.

2.2. The cost-effectiveness appraisal as a diagnostic test

There are three appraisal options in broad terms for appraisal organisations to choose. One extreme is to simply trust the information put forward by the producers and to approve all technologies with a claimed-ICER below λ . Given producers' incentives to claim ICERs below λ , this approach would likely lead to all new technologies being deemed cost-effective and approved. The other extreme option is to distrust producer submissions completely and to reject them all. Of course the latter would be very unpopular and political pressure or legislation are unlikely to allow for this. Both extremes are highly inaccurate, but low cost appraisal processes. A third alternative is trying to deduce which of the submitted technologies are actually cost-effective and which are not cost-effective. This third approach taken by most, if not all, appraisal organisations can be characterised as a diagnostic test aimed at identifying cost-effective technologies. Figure 2 outlines the outcomes and payoffs of having a diagnostic test.

Figure 2 about here

2.2.1. Outcomes

As seen in Figure 2, the outcome of the test can either be positive (Test+), i.e. technologies are deemed cost-effective, or negative (Test-), i.e. deemed to be not cost-effective. A positive test outcome can either be a *true-positive* (TP), i.e. the technology is deemed cost-effective and has a true-ICER below λ ; or a *false-positive* (FP), i.e. the technology is deemed cost-effective but has a true-ICER above λ . True-positives are cost-effective technologies that are correctly identified as being cost-effective, whereas false-positives are technologies that are identified as

How to design the cost-effectiveness appraisal process

being cost-effective but are actually not cost-effective. Similarly, a negative test outcome can be either *true-negative* (TN), i.e. the technology is deemed to be not cost-effective and has a true-ICER above λ ; or a *false-negative* (FN), i.e. the technology is deemed to be not cost-effective but has a true-ICER below λ .

The accuracy of diagnostic tests in terms of identifying true-positives and true-negatives is commonly presented in terms of sensitivity and specificity, defined as (Pagano & Gauvreau, 1993):

Sensitivity = p = true positive rate, i.e. probability of classifying a cost-effective technology as cost-effective

$1-p$ = false-negative rate, i.e. probability classifying a cost-effective technology as non-cost-effective

Specificity = q = true negative rate i.e. probability of classifying a non-cost-effective technology as non-cost-effective

$1-q$ = false-positive rate, i.e. probability of classifying a non-cost-effective technology as cost-effective

How an appraisal process is operated in terms of sensitivity and specificity determines the outcome and payoff from utilising the appraisal process, which is demonstrated in Section 3.

2.2.2. Payoffs

Figure 2 displays the payoffs to population health and producer surplus from having a diagnostic test. The payoff functions demonstrate that the gains from operating an appraisal

How to design the cost-effectiveness appraisal process

process are dependent on (in addition to the accuracy of the test) incremental net benefit (INB)³ per patient; time of relevance, i.e. time of reimbursement approval (t_τ) until loss of relevance (T); and the incidence of the disease being treated (I_t), which, combined, make up population INB⁴.

The payoff functions also show that the cost of operating an appraisal process is comprised of two parts: i) a monetary cost of performing the test (C_{Agency} and $C_{Producer}$); and ii) a time-related cost equal to the health and producer surplus foregone due to later implementation at time t_τ instead of time of availability (t_0).

Prices set by producers determine payoffs, since prices determine ΔC and thus INB. Hence, the optimal appraisal process is dependent on the prices set by producers, which is analysed further in Section 3.

2.3. Sensitivity and specificity in cost-effectiveness appraisal

In relation to cost-effectiveness assessment, sensitivity and specificity describe the ability of the appraisal process correctly to identify cost-effective (true-positive) and non-cost-effective (true-negative) technologies. Sensitivity and specificity are likely to be related to the true-ICER,

³ Incremental net benefit can be defined as either incremental net health benefits (INHB) or incremental net monetary benefits (INMB) in the following way (Stinnett & Mullahy, 1998; Tambour et al., 1998):

$$INHB = \Delta E - \frac{\Delta C}{\lambda} \text{ and } INMB = \Delta E * \lambda - \Delta C$$

⁴ Population incremental net benefit ($INB_{population}$) is based on aggregation of patient incremental net benefit ($INB_{patient}$) for the time of relevance, i.e. from reimbursement approval (t_τ) until loss of relevance (T), and number of patients, given the incidence (I_t), for which the technology is relevant:

$$INB_{population} = INB_{patient} * \sum_{t=t_\tau}^T \frac{I_t}{(1+r)^t}$$

How to design the cost-effectiveness appraisal process

since the further the true-ICER is above or below λ , the higher the probability is for the appraisal organisation to come to the correct conclusion, *ceteris paribus*. This is exemplified in Figure 3 below, which presents the probability of a positive cost-effectiveness appraisal outcome ($p[\text{test+}]$) as a function of the true-ICER ($p[\text{test+}]\text{-curve}$) based on an exemplified appraisal process $\alpha 1$ specified in Appendix 1.

Figure 3 about here

Technologies in the area below the curve (A and C) are considered to be cost-effective and those above the curve (B and D) are considered non-cost-effective. Of the technologies below the curve, those in area A are truly cost-effective (true-positives), whereas those in area C are non-cost-effective (false-positives). Similarly, technologies not considered cost-effective in area B are actually cost-effective (false-negatives), whereas those in area D are truly non-cost-effective (true-negatives).

There are several reasons why an appraisal process might come to the wrong conclusion regarding the cost-effectiveness of technologies. One example of technologies located in B (false-negatives) is cost-effective technologies (true-ICER $< \lambda$) with an underestimated treatment effect that results in an appraisal-ICER above λ . Another example is technologies that are considered non-cost-effective due to high uncertainty around the ICER due to, for example, uncertainty or limited evidence about effect or cost, but where further research shows that the true-ICER is below λ .

Sensitivity is equal to $\frac{A}{A+B}$, and specificity is equal to $\frac{D}{C+D}$ ⁵. Figure 3, as well as the following analysis, assumes that the probability of positive cost-effectiveness appraisal outcome

⁵ To be able to specify specificity (and 1-specificity) as $\frac{D}{C+D}$ it is necessary to define some upper bound for D. Otherwise D would be boundless and $\lim_{\text{true-ICER} \rightarrow \infty} \left(\frac{D}{C+D}\right) = 1$ and $\lim_{\text{true-ICER} \rightarrow \infty} \left(\frac{C}{C+D}\right) = 0$, regardless of the size of C. In the following numerical examples the upper limit is defined as $3*\lambda$, but it could just as well be set at $10*\lambda$. Setting an upper limit does not change the accuracy of the process only how accuracy is calculated and reported.

How to design the cost-effectiveness appraisal process

depends mainly on the true-ICER and, to a lesser extent, on factors such as, for example, the claimed-ICER.

2.3.1. Accuracy vs. cost and sensitivity vs. specificity trade-off.

Appraisal organisations can increase sensitivity and specificity, i.e. $\frac{A}{A+B}$ and $\frac{D}{C+D}$, by operating a more accurate appraisal process. This is exemplified by the p[test+]-curve for the more accurate appraisal process β_1 shown in Figure 4.a. Increased accuracy could be achieved by waiting until more relevant evidence becomes available or increasing the level of scrutiny, i.e. doing more detailed assessment, further modelling or additional literature reviews. Ideally, appraisal bodies would like to have a (zero cost) test with perfect accuracy, i.e. with sensitivity and specificity equal to 1⁶. However, improving the accuracy of the appraisal process comes at the expense of a higher monetary (C_{Agency}) and/or time cost (t_0-t_τ), whereby there is a trade-off between the accuracy and the cost of operating an appraisal process.

Figure 4 about here

Appraisal organisations can also choose to increase sensitivity or specificity without increasing scrutiny or otherwise changing the process. This can be done by changing the definition of what is considered to be cost-effective. Appraisal organisations can do this by allowing for more (or less) uncertainty around ICER estimates. This could, for example, be done by including the outcome from probabilistic sensitivity analyses (PSA) or value of information analysis (VOI) in the cost-effectiveness assessment (Griffin et al., 2011; Claxton, 1999). Setting specific limits for the percentage of simulations in PSA that are required to be below λ ; setting limits for the maximum proportion of simulated ICERs that has “high” values, e.g. above $2x\lambda$; or setting

⁶ In contrast, the producer would prefer a process with sensitivity equal to 1 and specificity equal to 0, which is equivalent to approving all technologies. This demonstrates how appraisal organisations and producers have different incentives in terms of how the appraisal process should be designed.

How to design the cost-effectiveness appraisal process

maximum levels of VOI that will be accepted, are different ways that appraisal organisations can set, and change, their definition of what is considered cost-effective⁷.

Accepting more uncertainty around the ICER would lead to more technologies being considered cost-effective and thus increase sensitivity ($\frac{A}{A+B}$) but reduce specificity ($\frac{D}{C+D}$). This is exemplified by curve α_2 in Figure 4.b. Similarly, appraisal organisations can implement more strict criteria for what is considered cost-effective leading to more technologies being determined as non-cost-effective, i.e. decreasing sensitivity and increasing specificity, as exemplified by curve α_3 in Figure 4.b.

The relationship between sensitivity and specificity within diagnostic tests is often expressed in a Receiver Operating Characteristic (ROC) curve (Pagano & Gauvreau, 1993). Figure 4.c represents the ROC-curve for the appraisal process α and the more accurate process β . The shape of ROC-curves is determined by the accuracy of the test. Tests that are better than pure chance will have a concave shape and the ROC-curve for a more accurate test will lie above the curve of a less accurate test.

A p[test+]-curve represents one specific sensitivity and specificity combination and thus one point on a ROC-curve. The points α_1 , α_2 , and α_3 on ROC-curve α in Figure 4.c correspond to the different p[test+]-curves α_1 , α_2 , and α_3 in Figure 4.b. Moving along the ROC-curve represents different sensitivity and specificity combinations and hence different p[test+]-curves that the appraisal organisation can choose between when operating a given appraisal process.

⁷ For example, if appraisal organisations require that $\geq 75\%$ of PSA simulations are $< \lambda$, then more technologies will be deemed not-cost-effective compared to a requirement of $\geq 50\%$ of simulated ICERs being $< \lambda$. Similarly, requiring that only 5% of simulated ICERs can be $> 2x\lambda$ will lead to fewer technologies being deemed cost-effective, compared to allowing for more simulated ICERs estimates being $> 2x\lambda$. Setting low maximum VOI limits that will be accepted will, likewise, lead to more technologies being deemed not-cost-effective, compared to allowing for higher VOI estimates.

How to design the cost-effectiveness appraisal process

Defining the cost-effectiveness appraisal as a diagnostic test demonstrates that finding the optimal design of the cost-effectiveness appraisal process is a trade-off between accuracy versus cost, i.e. finding the optimal ROC-curve shape. It can further be concluded that the question of how to best operate a given appraisal process, with certain time and resource constraints, is a matter of finding the optimal sensitivity and specificity combination, i.e. determining the optimal point on a given ROC-curve. In the following section we will use this framework to analyse how best to design and to operate the appraisal process in order to maximise population health.

3. How to design and to operate the cost-effectiveness appraisal process

In this section, we analyse how best to design the appraisal process, under the assumption that producers' price setting is exogenous to the appraisal process. The assumption about exogenous price setting will be relaxed in the final part of the analysis, which also assesses how allowing for appeals, pricing agreements, resubmissions and re-evaluations impacts optimal design.

3.1. How to operate a specific appraisal process to maximise population health

Getting the most health from utilising a given appraisal process is a question of finding the optimal sensitivity and specificity combination. However, the optimal sensitivity and specificity combination is dependent on the INB of future technologies, since the value of correctly identifying cost-effective and non-cost-effective technologies depends on the number and INB of future technologies. Figure 5 presents a stylised example demonstrating how the optimal way to operate the appraisal process is endogenous to the population INB of future technologies and thus to prices set by producers. In this example the process α would be operated optimally, in terms of population health, at sensitivity and specificity combination α^* and α^{**} , given ICER distribution I and II, respectively.

Figure 5 about here

How to design the cost-effectiveness appraisal process

In general it can be said that the optimal point will move upwards and to the right along the ROC-curve (increasing sensitivity and decreasing specificity) when the value of identifying cost-effective technologies increases, i.e. when there is a higher proportion, larger patient population (I_t) or time relevance ($T-\tau$) for cost-effective technologies vs. non-cost-effective technologies, and vice versa. Hence, there is no one optimal way to operate an appraisal process in terms of sensitivity and specificity.

3.1.1. Burden of Proof

An issue that can restrict the way in which an appraisal process can be operated is requirements for appraisal organisations to provide proof of technologies' cost-effectiveness, i.e. the burden of proof. If appraisal organisations are obligated to ensure that cost-effective technologies are identified, then they must operate the process towards the top (right) end of ROC-curve with high sensitivity; because only then can appraisal organisations ensure that (most) cost-effective technologies are identified. Similarly, if appraisal organisations are obligated to ensure that non-cost-effective technologies are identified, then the process has to be operated on the left (and bottom) side of the ROC-curve.

In either case, placing the burden of proof on the appraisal organisation will restrict the way in which the process can be operated and is likely to result in suboptimal operation of the appraisal process. In contrast, if it is up to producers to provide proof of cost-effectiveness then appraisal organisations are free to choose the (optimal) sensitivity and specificity combination.

3.2. Optimal design of an appraisal process

As previously discussed, population health benefits from operating an appraisal process are improved by increasing accuracy. However, a more accurate process comes at a higher monetary (C_{Agency}) and/or time cost (t_0-t_τ), whereby the gains of a more accurate process must be weight against the increased cost.

How to design the cost-effectiveness appraisal process

Improving the accuracy will reduce the NB forgone from false-negatives and lost from false-positives. The opportunity cost of false-negatives and the cost of false-positives is dependent on the population INBs of the technologies that are incorrectly rejected and accepted. Hence, the value of increased accuracy is dependent on the population INBs and proportion of future cost-effective and non-cost-effective technologies. Since the value of improving accuracy is dependent on the population INBs of future technologies, there is no one optimal level of accuracy.

Still some insight can be gained into when it might be valuable to trade-off time and money for a more accurate process. The impact of false-positives and false-negatives increases when the population INBs of these technologies are large. Hence, the value of improving accuracy increases when: i) technologies have larger patient populations and/or longer time of relevance, which increase the loss from incorrect decisions; and ii) there is larger uncertainty around the ICER estimate leading to higher likelihood of incorrect decisions. Increasing accuracy can reduce the likely inefficiency of operating the appraisal process with a specific level of sensitivity or specificity; thereby increasing the value of operating a more rigorous cost-effectiveness appraisal when burden of proof is placed on appraisal organisations.

3.3. Pricing of future technologies might be endogenous to the appraisal process

The previous sections have demonstrated that optimal design and operation of an appraisal process is dependent on the population INB of future technologies. There is reason to believe that the design and operation of the appraisal process might influence the price setting of future technologies, whereby the INB of future technologies might be endogenous to the appraisal process.

How to design the cost-effectiveness appraisal process

Technologies with lower true-ICERs have higher probability of being deemed cost-effective compared to technologies with higher true-ICERs. This implies that producers face a trade-off between setting a higher price (and ICER) to collect higher producer surplus if approved and setting a lower price (and ICER) to increase the probability of being deemed cost-effective. Since the design and operation of the appraisal process determines the probability of positive cost-effectiveness appraisal, i.e. the $p[\text{test+}]$ -curves, changing the design or operation of the process may change pricing incentives. This is exemplified in Figure 6 which utilises the numerical example to present expected producer surplus ($E[\text{Producer surplus}]$) as a function of the true-ICER for the two different $p[\text{test+}]$ -curves α_2 and α_3 .

Figure 6 about here

In this example, operating the test at a lower point on the ROC-curve changes producers' incentives from pricing above λ to pricing below λ , assuming that producers will set the price to maximise expected producer surplus. However, the lower sensitivity from operating at a lower point on the ROC-curve also reduces the ability to detect and accept truly cost-effective technologies. Appraisal organisations must, therefore, find the point on the ROC-curve that balances the trade-off between incentivising producers to put forward cost-effective technologies and the ability to detect and accept cost-effective technologies.

Improving accuracy and thereby sensitivity and/or specificity increases producers' incentives to price below λ since this increases the expected producer surplus of pricing below λ and/or reduces the expected producer surplus of pricing above λ . Therefore, when producer price setting is endogenous to the appraisal process, the gain from improving accuracy is twofold: i) correctly identifying more truly cost-effective and non-cost-effective technologies; and ii) increasing incentives for cost-effective price setting. Still, the benefits of improving accuracy must be weighed against the cost from operating a more rigorous process.

How to design the cost-effectiveness appraisal process

3.4. The appraisal process may not be a “one-off test”

The analysis has so far characterised the appraisal process as a one-off test. In this section we relax this assumption and analyse how appeals, price agreements, resubmissions and re-evaluations influences optimal design of the appraisal process.

3.4.1. Appeals

Having an appeal process can reduce the loss from false-negatives and false-positives. Given the informational advantage that producers hold and the cost of appealing, producers are probably more likely to appeal false-negatives compared to payers appealing false-positives. Hence, introducing appeals will probably reduce the loss from false-negatives more than the loss from false-positives making it optimal to increase the proportion of false-negatives compared to false-positives, i.e. operating the appraisal process on a lower point on the ROC-curve with higher specificity and lower sensitivity.

Producers may also appeal true-negatives to have a second go at getting non-cost-effective technologies approved. All this should be taken into consideration when designing the appeal process, even if the main benefit of an appeal process might be the procedural value and to avoid long and expensive court cases.

3.4.2. Price agreements and resubmissions

It is important to distinguish between three types of price agreements: pre-, during- and post-appraisal agreements. Pre-appraisal agreements are basically just a lowering of the (effective) price of a new technology. During-appraisal price reductions allow producers to adjust the price during the appraisal, which producers may choose to do if they are informed that the technology will be rejected at the current price. Post-appraisal price reductions could be used if producers get a negative appraisal and are able to resubmit for a new appraisal with a lower price.

How to design the cost-effectiveness appraisal process

Allowing for during- and post-appraisal agreements can, therefore, turn true-negatives and false-negatives into true-positives (and false-positives).

If price setting is endogenous to the appraisal process then allowing during- and post-appraisal agreements could lead to negative pricing incentives, i.e. incentivise producers to set higher prices and put forward more non-cost-effective technologies and then adjust the price during the process. Further, if the appraisal organisation has overestimated ΔE or underestimated ΔC it may lead to false-positives where producers still charge a price that results in true-ICERs above λ .

Appraisal organisations can decrease the negative pricing incentives of appeals, pricing agreements and resubmissions by making it more costly to utilise these options. Not allowing during-appraisal price reduction but only for resubmissions with lower price or otherwise increasing the monetary and/or the time cost of these options, are some options to increase the cost to producers.

3.4.3. Re-evaluations

Many appraisal organisations utilise conditional approvals and re-evaluations (Stafinski et al., 2010). These options enable review and reversal of past decisions, which can reduce the loss from false-positives. With a lower cost of false-positives it will be optimal to accept more technologies, i.e. operating on a higher point on the ROC curve with higher sensitivity and lower specificity. The framework indicates that the best candidates for re-evaluations will be technologies with large patient populations, long time of relevance and high uncertainty around the ICER, since these technologies have high consequence or risk of incorrect decisions.

How to design the cost-effectiveness appraisal process

4. Discussion

The framework outlined in this paper demonstrates that there is no *one* optimal way to design and to operate the cost-effectiveness appraisal of new health care technologies. The trade-off between accuracy and cost of the appraisal process implies that even a (hypothetical) process with 100% accuracy is unlikely to be optimal, assuming decreasing marginal returns of improving accuracy. Optimal design and operation depend on the population INB and thereby price, size of patient population and the time of relevance of the technologies that will undergo assessment. Optimal design will, therefore, change over time, whereby it is important that the operational and legal framework ensures continuous evaluation and adjustment of the cost-effectiveness appraisal process over time.

Our results further indicate that it may be valuable to adjust the process according to the specific technology being assessed. A ‘one size fits all’ appraisal process is likely to be inefficient, and may be improved by a flexible approach directing time and resources to appraisals where they are expected to provide most value. For example, a more accurate process may be warranted when technologies have large patient populations or will be used for many years, since these factors increase the cost of incorrect decisions, and vice versa. This may justify the recent changes to the NICE process, i.e. spending less time and resources in the “fast track” appraisal of technologies with an expected ICER below £10,000 per QALY that have a limited budget impact; but spending additional time and resources on “commercial discussions” between producers and the NHS for technologies with significant budget impact, i.e. above £20m (NICE, 2017a; NICE, 2017b). However, much depends on the detail of how such arrangements will be implemented. In particular, how an expected ICER below £10,000 will be verified and the additional accuracy for ‘high priority’ appraisals will be delivered.

How to design the cost-effectiveness appraisal process

If pricing of technologies is exogenous to the appraisal process, our analysis shows that optimal design and operation is a two-part optimisation problem of accuracy versus cost and sensitivity versus specificity. The first part relates to selecting an appraisal process design (which determines the shape of the ROC-curve in the conceptual framework); and the second part, considers the optimal way to operate a given process (finding the optimal point on a particular ROC-curve in the conceptual framework). If pricing is endogenous to the appraisal process the optimal design of the process becomes more complex and may be characterised as a game, where appraisal organisations need also to consider producers' reactions to the design and operation of the appraisal process. The widespread use of international reference pricing, and focus on price differences across countries, makes price reductions due to single country appraisal design less likely, at least in European countries given potential price and revenue spill-over to other countries (Kanavos et al., 2010; Danzon et al., 2005). However, the ability of producers to offer country-specific and confidential discounts, and the fact that decisions made by some appraisal bodies, like for example NICE, may influence other appraisal bodies decisions, increase the likelihood of producers adjusting prices to increase the probability of approval (Sculpher, 2010; Kennedy, 2009). Assessment of the endogeneity between the appraisal process and producers' price setting, therefore, needs to be considered on a country level when designing of the appraisal process.

The endogeneity between price setting and the appraisal process and the determinants of population INB (size of the patient population, current clinical practice, and price setting) will vary between countries. Our framework therefore indicates that the optimal appraisal process ought to differ between countries, which may explain the observed variation in appraisal processes across countries. For example, the larger population of the UK and the fact that NICE's decisions are referenced by other countries could explain why the NICE process has been longer and includes the extensive external ERG review, compared to the shorter and most

How to design the cost-effectiveness appraisal process

likely less rigorous SMC and TLV processes in Scotland and Sweden, respectively. However, research is needed to evaluate if differences in appraisal processes across countries can be explained by different assessments of the optimal sensitivity versus specificity and accuracy versus cost trade-offs; or whether other factors explain observed differences in cost-effectiveness appraisal processes.

Many countries allow producers to propose during- and post-appraisal price reductions, for example, in the form of confidential discounts or patient access schemes (Kanavos et al., 2011; Kanavos et al., 2010). These are attractive policies since they enable the approval of technologies that would otherwise have been deemed non-cost-effective and rejected. However, our analysis demonstrates how these price agreements may increase producers' incentives to price above the threshold. Authorities should, therefore, consider how they can minimise these adverse pricing incentives when discounts or patient access schemes are part of the process.

The framework underlines an important distinction regarding handling of uncertainty in cost-effectiveness assessments: i) reducing uncertainty, i.e. accuracy versus cost; and ii) how to take uncertainty into account, i.e. sensitivity versus specificity. Given time and resource constraints, appraisal organisations seldom have the possibility of reducing uncertainty (for example, through commissioning additional research) and, instead, need to decide on how to take uncertainty into account (Griffin et al., 2011). Sensitivity analysis, and the display of uncertainty using cost-effectiveness acceptability curves/frontiers, plays a key role in quantifying uncertainty in cost-effectiveness appraisals (Heintz et al., 2016; Claxton, 2008). Even though these are standard parts of most cost-effectiveness assessments today, the way in which these are used as decision criteria, and ultimately determine the proportion of false-positive and false negatives, needs further clarification and consideration, informed by recent work in this area (Claxton et al., 2012; Claxton et al., 2016; Walker et al., 2012).

How to design the cost-effectiveness appraisal process

There is increasing focus on “fast access to medicines” (European Medicines Agency, 2015b; European Medicines Agency, 2015a; Accelerated Access, 2015; U.S. Department of Health and Human Services et al., 2014). Given the trade-off between accuracy and time/cost, our analysis indicate that the move toward faster access is likely to increase the number of false-positive and/or false-negatives. It would, therefore, be relevant to study the population health effect of speeding up the access to new medicines.

Our framework assumes that cost-effectiveness assessment is based solely on the expected/appraisal ICER and the cost-effectiveness threshold. Even though cost-effectiveness has been shown to be the strongest predictor of the reimbursement recommendations made by NICE (Dakin et al., 2015), this is naturally a simplification of the appraisal process. The framework can be extended to incorporate different cost-effectiveness thresholds, as for example employed by NICE with the end-of-life criteria (Paulden et al., 2014; Stewart et al., 2014), or threshold as a function of, for example, disease severity and/or medical need, as appear to be employed in for instance Sweden and the Netherlands (Franken et al., 2014; Franken et al., 2012; Liliemark et al., 2016). Nevertheless, the conclusions and policy implications drawn from the simple framework presented in this paper are expected to hold, even when increasing the complexity with several or varying thresholds.

In conclusion, based on our analysis we draw some important policy conclusions: i) there is no *one* optimal appraisal process and the appraisal should be adapted over time and to the specific technology under assessment; ii) how price reductions and patient access schemes impact producer’s price setting has to be considered when designing the appraisal process; iii) burden of proof should be on the producers; and iv) PSA and VOI decision rules can be used to adjust the sensitivity and specificity combination of the appraisal process. This underlines the need for further and continued analysis of how to design and adapt the appraisal process over time.

How to design the cost-effectiveness appraisal process

5. References

- Accelerated Access, 2015. Accelerated Access Review: Interim report.
- Barbieri, M., Hawkins, N. & Sculpher, M., 2009. Who does the numbers? The role of third-party technology assessment to inform health systems' decision-making about the funding of health technologies. *Value in Health* **12**(2): 193–201.
- Barham, L., 2008. Single technology appraisals by NICE: Are they delivering faster guidance to the NHS? *Pharmacoeconomics* **26**(12): 1037–1043.
- Bell, C.M. et al., 2006. Bias in published cost effectiveness studies: systematic review. *BMJ* **332**(7543): 699–703.
- Bell, J., 2015. The accelerated access review, my reflections and the case for change. In *Accelerated Access Review: Interim report*.
- Briggs, A., Sculpher, M. & Claxton, K., 2006. *Decision modelling for health economic evaluation*. First edition, Oxford university press: Oxford.
- Chauhan, D., Miners, A.H. & Fischer, A.J., 2007. Exploration of the difference in results of economic submissions to the National Institute of Clinical Excellence by manufacturers and assessment groups. *International journal of technology assessment in health care* **23**(1): 96–100.
- Claxton, K., 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of health economics* **18**(3): 341–64.
- Claxton, K., 2008. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics* **26**(9): 781–798.
- Claxton, K. et al., 2010. *Appropriate Perspectives for Health Care Decisions*. *CHE Research Paper 54*.
- Claxton, K. et al., 2012. Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme

How to design the cost-effectiveness appraisal process

- of evidence development. *Health Technology Assessment* **16**(46).
- Claxton, K. et al., 2015. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technology Assessment* **19**(14): 1–504.
- Claxton, K. et al., 2016. A Comprehensive Algorithm for Approval of Health Technologies With , Without , or Only in Research: The Key Principles for Informing Coverage Decisions. *Value in Health* **19**: 885–891.
- Dakin, H. et al., 2015. The influence of cost-effectiveness and other factors on nice decisions. *Health economics* **24**: 1256–1271.
- Danzon, P.M., Wang, Y.R. & Wang, L., 2005. The impact of price regulation on the launch delay of new drugs--evidence from twenty-five major markets in the 1990s. *Health economics* **14**(3): 269–92.
- 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**(5): 437–52.
- Drummond, M.F. et al., 2015. *Methods for the Economic Evaluation of Health Care Programmes*. Fourth edi., Oxford University Press: Oxford.
- European Medicines Agency, 2015a. Adaptive pathways. http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000601.jsp [16 May 2016].
- European Medicines Agency, 2015b. Fast track routes for medicines that address unmet medical needs. http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2015/07/news_detail_002381.jsp&mid=WC0b01ac058004d5c1 [16 May 2016].
- Franken, M. et al., 2012. Similarities and differences between five European drug

How to design the cost-effectiveness appraisal process

- reimbursement systems. *International Journal of Technology Assessment in Health Care* **28**(4): 349–357.
- Franken, M., Koopmanschap, M. & Steenhoek, A., 2014. Health economic evaluations in reimbursement decision making in the Netherlands : Time to take it seriously ?
- George, B., Harris, A. & Mitchell, A., 2001. Cost-effectiveness analysis and the consistency of decision making: evidence from pharmaceutical reimbursement in Australia (1991 to 1996). *Pharmacoeconomics* **19**(11): 1103–9.
- Griffin, S.C. et al., 2011. Dangerous omissions : the consequences of ignoring decision uncertainty. *Health Economics* **224**(20): 212–224.
- Heintz, E., Salah, A.G. & Francoise, G., 2016. Is There a European View on Health Economic Evaluations ? Results from a Synopsis of Methodological Guidelines Used in the EUnetHTA Partner Countries. *Pharmacoeconomics* **34**(1): 59–76.
- Kaltenthaler, E. et al., 2008. Comparing methods for full versus single technology appraisal: A case study of docetaxel and paclitaxel for early breast cancer. *Health Policy* **87**(3): 389–400.
- Kanavos, P. et al., 2011. *Differences in costs of and access to pharmaceutical products in the EU. European Parliament.*
- Kanavos, P. et al., 2010. *Short- and Long-Term Effects of Value-Based Pricing vs . External Price Referencing.*
- Kennedy, I., 2009. *Appraising the value of innovation and other benefits, A short study for NICE.*
- Liliemark, J. et al., 2016. Betalningsviljan för nya läkemedel bygger på etiska principer. *Läkartidningen* **113**: 1–5.
- McCabe, C.J. et al., 2010. Access with evidence development schemes: A framework for description and evaluation. *Pharmacoeconomics* **28**(2): 143–152.

How to design the cost-effectiveness appraisal process

- McKenna, C. et al., 2015. Unifying Research and Reimbursement Decisions: Case Studies Demonstrating the Sequence of Assessment and Judgments Required. *Value in Health* **18**(6): 865–875.
- Miners, A.H. et al., 2005. Comparing estimates of cost effectiveness submitted to the National Institute for Clinical Excellence (NICE) by different organisations: retrospective study. *BMJ* **330**(65): 1–4.
- NICE, 2014. *Guide to the processes of technology appraisal*.
- NICE, 2017a. *Fast track appraisal - Addendum to the Guide to process of technology appraisal*.
- NICE, 2017b. *Technology Appraisal and Highly Specialised Technologies Programmes - Procedure for varying the funding requirement to take account of net budget impact*.
- Pagano, M. & Gauvreau, K., 1993. *Principles of Biostatistics.*, Duxbury Press: Belmont, California.
- Paulden, M. et al., 2014. Some Inconsistencies in NICE 's Consideration of Social Values. *Pharma* **32**: 1043–1053.
- Rawlins, M.D. & Culyer, A.J., 2004. National Institute for Clinical Excellence and its value judgments. *BMJ* **329**: 224–227.
- Sculpher, M., 2010. Single technology appraisal at the UK National Institute for Health and clinical excellence: a source of evidence and analysis for decision making internationally. *Pharmacoeconomics* **28**(5): 347–349.
- Stafinski, T., McCabe, C.J. & Menon, D., 2010. Funding the Unfundable. **28**(2): 113–142.
- Stewart, G. et al., 2014. The Impact Of Nice's End-Of-Life Threshold On Patient Access To New Cancer Therapies In England And Wales. In *ISPOR 19th Annual International Meeting*.
- Stinnett, A.A. & Mullahy, J., 1998. Net Health Benefits : A New Framework for the Analysis of Cost-Effectiveness Analysis Uncertainty in. *Medical Decision Making* **18**(2)

How to design the cost-effectiveness appraisal process

Supplement): S68-80.

Tambour, M., Zethraeus, N. & Johannesson, M., 1998. A note on confidence intervals in cost-effectiveness analysis. *Int J Technol Assess Health Care* **14**(3): 467–71.

TLV, 2016. Processing time. <http://tlv.se/In-English/medicines-new/apply-for-a-price-or-reimbursement/processing/> [16 May 2016].

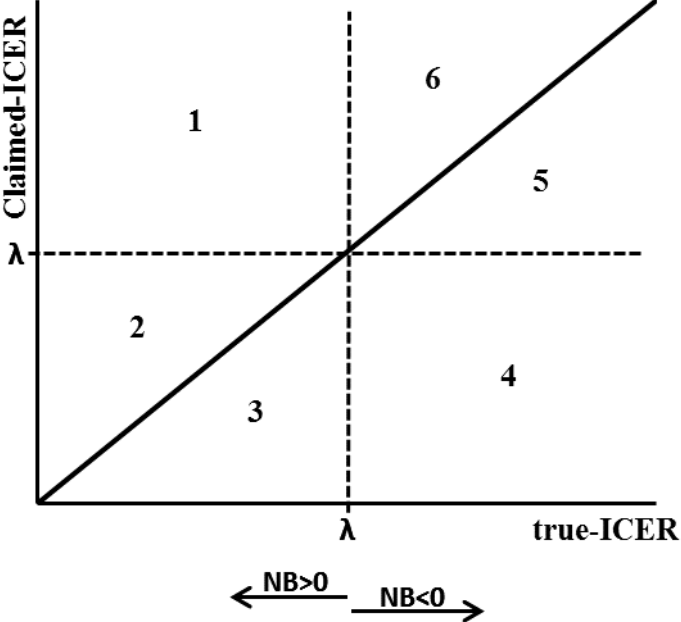
U.S. Department of Health and Human Services et al., 2014. *Guidance for Industry: Expedited Programs for Serious Conditions – Drugs and Biologics*.

Walker, S. et al., 2012. Coverage with evidence development, only in research, risk sharing, or patient access scheme? a framework for coverage decisions. *Value in Health* **15**(3): 570–579.

How to design the cost-effectiveness appraisal process

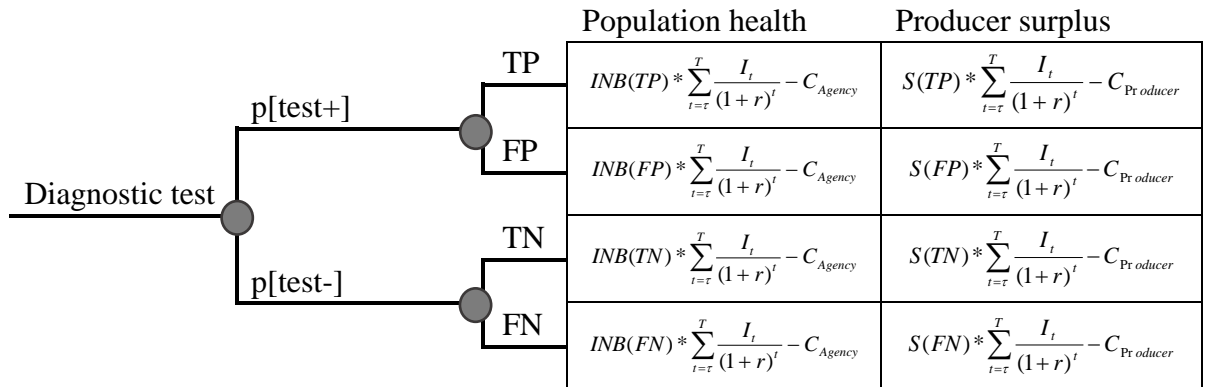
Figures

Figure 1. Possible combinations of claimed- and true-ICERs



How to design the cost-effectiveness appraisal process

Figure 2. Decision tree outlining the three different appraisal approaches and corresponding payoffs



Where:

P[Test+]/P[Test-] represents the probability of a positive/negative test outcome

TP/FP is true-/false-positives and TN/FN is true-/false-negatives

INB() and S() represents the incremental net benefit and producer surplus of the given outcome

I_t is the incidence in time period t

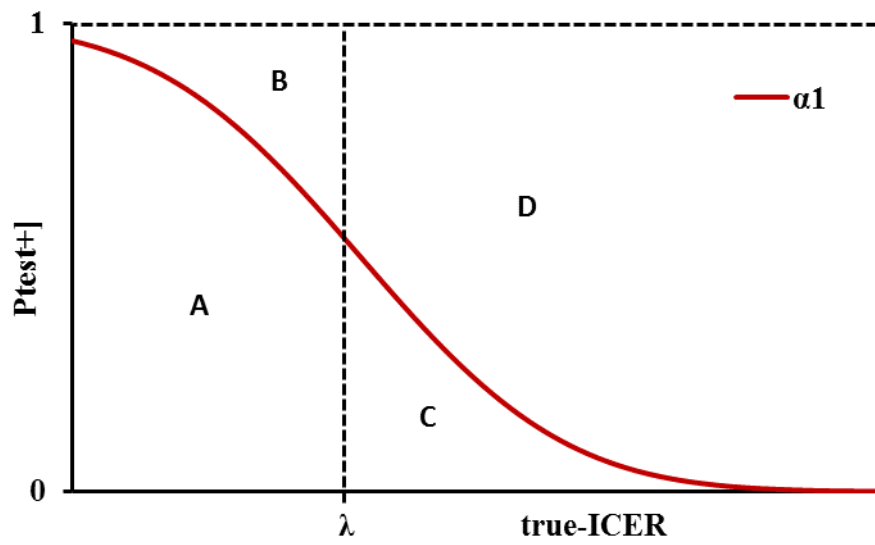
t_τ is the time that the appraisal decision is ready

T is the time when the technology losses relevance

r is the discount rate

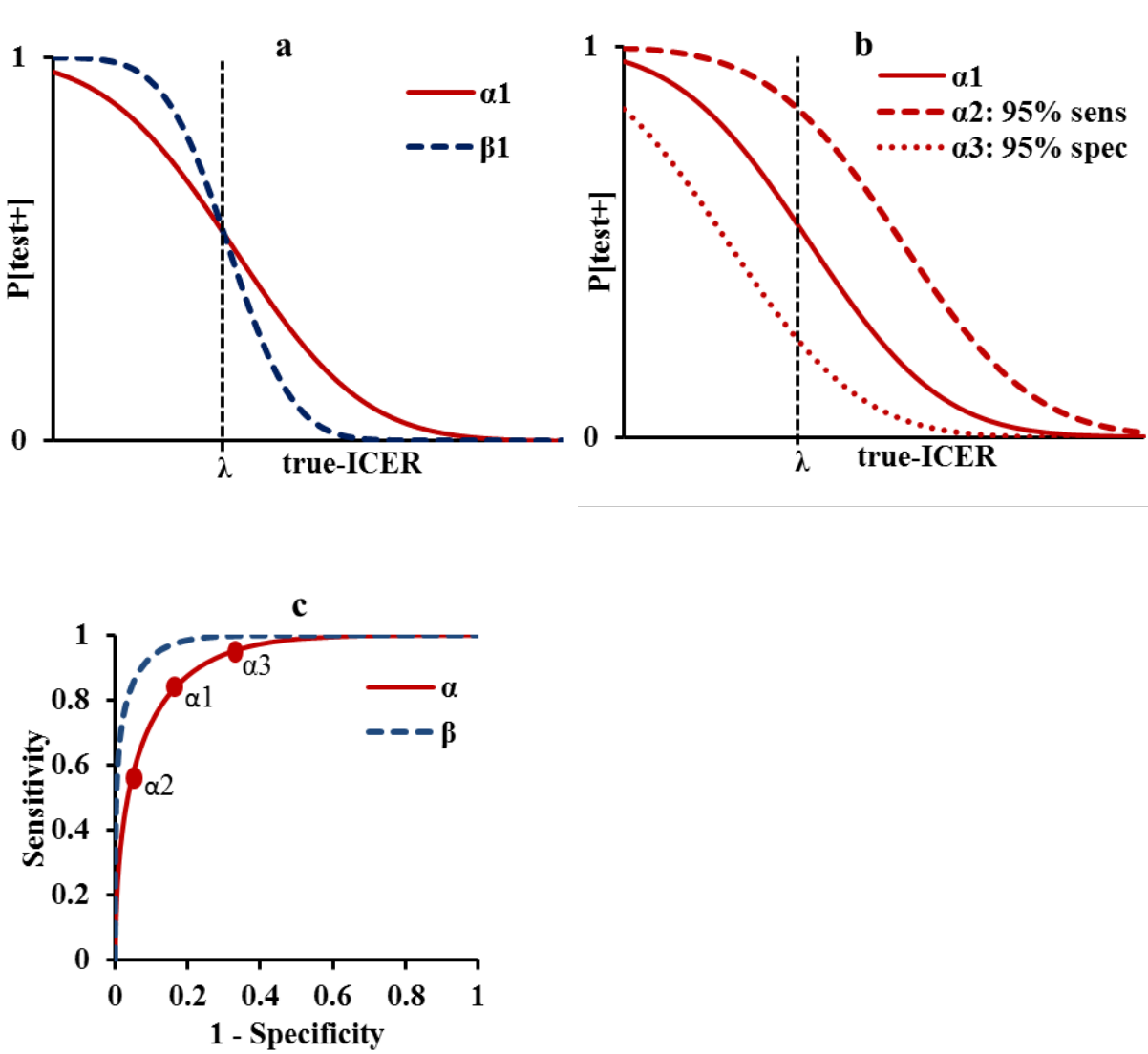
How to design the cost-effectiveness appraisal process

Figure 3. Probability of positive cost-effectiveness appraisal ($p[\text{test+}]$) as a function of the true-ICER (based on appraisal process α_1 in Appendix 1)



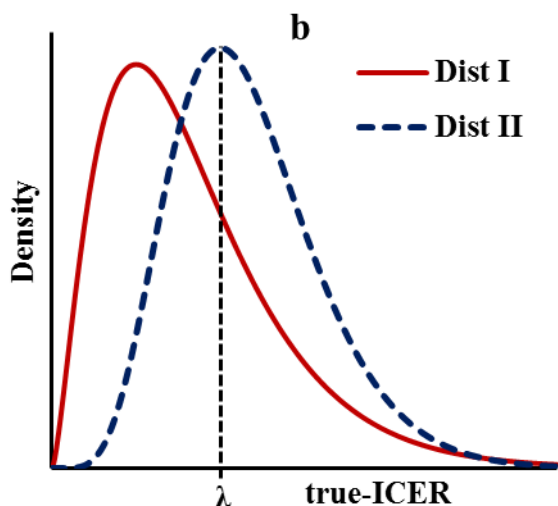
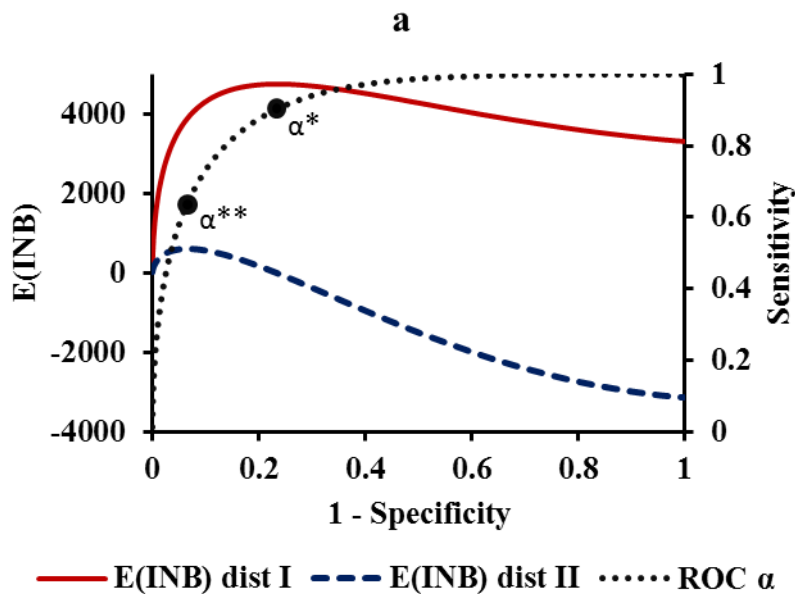
How to design the cost-effectiveness appraisal process

Figure 4. (a) Probability of positive cost-effectiveness appraisal as a function of the true-ICER for process α and β ; (b) Probability of positive cost-effectiveness appraisal as function of true-ICER for process α at different sensitivity and specificity combinations (α_1 , α_2 and α_3); (c) Receiver operating characteristic (ROC) curve for process α and β . (based on process α and β defined in Appendix 1)



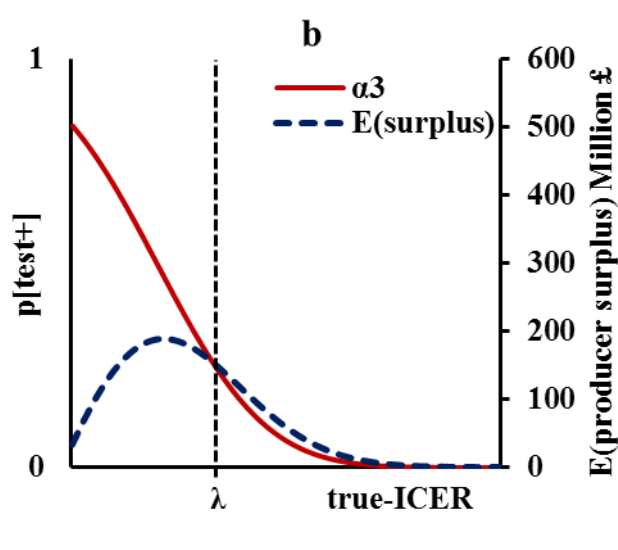
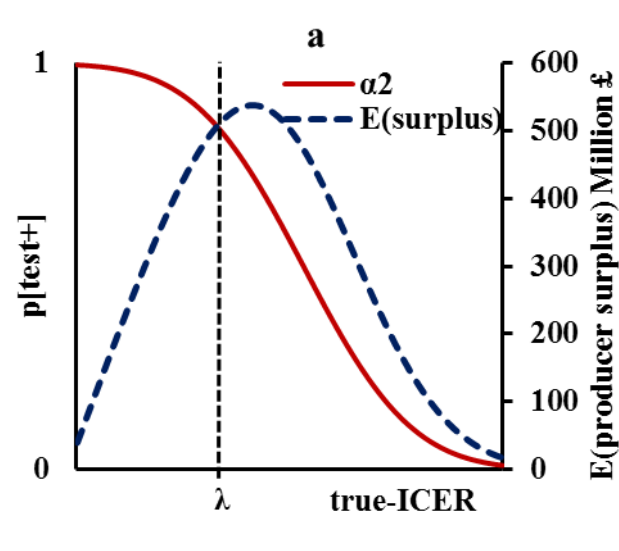
How to design the cost-effectiveness appraisal process

Figure 5. (a) Expected incremental net health benefits ($E(INB)$) as function of the sensitivity and specificity combinations of (the ROC-curve of) process α and ICER distributions I and II. The points α^* and α^{**} represents the sensitivity and specificity combinations where population health is maximised given ICER distributions I and II, respectively; (b) ICER distribution I and II. (based on the numerical example in Appendix 1)



How to design the cost-effectiveness appraisal process

Figure 6. (a) and (b) Expected producer surplus ($E(\text{producer surplus})$) as function of true-ICER given the $p[\text{test+}]$ -curves α_2 and α_3 , respectively. (based on process α and numerical example in Appendix 1)



How to design the cost-effectiveness appraisal process

Appendix 1

Characteristics of technologies	notation	value
Incremental QALY gain of the new technologies	ΔE	0.5
Production and distribution cost of the new technologies (£)	MC_{New}	100
Cost of using the current technology (£)	$C_{Alternative}$	1,000
Calculation of ΔC		$\Delta C = ICER * \Delta E$
Calculation of INB (INHB)		$INHB = \Delta E - \Delta C / \lambda$
Calculation of Producer Surplus		$\Delta C + C_{alternative} - MC_{New}$

The numerical example assumes that all new technologies have the same ΔE . The ΔC is estimated based on ΔE and an ICER drawn from the ICER distribution. When estimating producer surplus we assume that revenue gained from the new technology is equal to $\Delta C + C_{alternative}$. This assumes that there is no other relevant cost associated with using the new or the old technology. This was assumed to simplify the numerical example but potentially leads to an overestimation of producer surplus. However, this simplification has no effect on the interpretation of the numerical example or the general learnings from this framework.

Characteristics of the appraisal process	notation	value
Sensitivity	p	0.9
Specificity	q	0.7
Monetary cost of the appraisal process to the public (£)	C_{agency}	100,000
Monetary cost of the appraisal process to producers (£)	$C_{producer}$	150,000

Additional information needed	notation	value
Cost-effectiveness threshold (£/QALY)	λ	30,000
Yearly patient incidence/patient population	I	5,000
Discount rate (%)	r	3.50%
Life time of the new technologies (years)	T	10
Time point at which a decision is reached when using the appraisal process (years)	τ	1

How to design the cost-effectiveness appraisal process

ICER distributions	Ture-ICER distribution = Probability mass function
I	<i>Gamma(2.5; 10,000)</i>
II	<i>Gamma(7; 5,000)</i>

P[test+]-curves	P[test+] = 1-CDF (CDF is the cumulative density function)
Process α	<i>Normal(μ; 18,000)</i>
α_1	$\mu=35,000$
α_2	$\mu=48,000$
α_3	$\mu=18,000$
Process β	<i>Normal(μ; 9,000)</i>
When estimating sensitivity and specificity for process α and β an upper true-ICER limit is defined as $3*\lambda$. An upper limit is needed since specificity would otherwise be equal to or close to 1 regardless of the performance of the test, as described in footnote 3.	