



Severity-Adjusted Probability of Being Cost Effective

Matthijs M. Versteegh¹ · Isaac Corro Ramos¹ · Nasuh C. Buyukkaramikli¹ · Amir Ansaripour² · Vivian T. Reckers-Droog² · Werner B. F. Brouwer²

Published online: 27 May 2019
© The Author(s) 2019

Abstract

Background In the context of priority setting, a differential cost-effectiveness threshold can be used to reflect a higher societal willingness to pay for quality-adjusted life-year gains in the worse off. However, uncertainty in the estimate of severity can lead to problems when evaluating the outcomes of cost-effectiveness analyses.

Objectives This study standardizes the assessment of severity, integrates its uncertainty with the uncertainty in cost-effectiveness results and provides decision makers with a new estimate: the severity-adjusted probability of being cost effective.

Methods Severity is expressed in proportional and absolute shortfall and estimated using life tables and country-specific EQ-5D values. We use the three severity-based cost-effectiveness thresholds (€20,000, €50,000 and €80,000, per QALY) adopted in The Netherlands. We exemplify procedures of integrating uncertainty with a stylized example of a hypothetical oncology treatment.

Results Applying our methods, taking into account the uncertainty in the cost-effectiveness results and in the estimation of severity identifies the likelihood of an intervention being cost effective when there is uncertainty about the appropriate severity-based cost-effectiveness threshold.

Conclusions Higher willingness-to-pay thresholds for severe diseases are implemented in countries to reflect societal concerns for an equitable distribution of resources. However, the estimates of severity are uncertain, patient populations are heterogeneous, and this can be accounted for with the severity-adjusted probability of being cost effective proposed in this study. The application to the Netherlands suggests that not adopting the new method could result in incorrect decisions in the reimbursement of new health technologies.

Key Points for Decision Makers

Some societies have policies that indicate higher willingness to pay for the worse-off.

The Netherlands has formally introduced differential thresholds based on severity of illness of €20,000, €50,000 and €80,000 per QALY based on estimates of proportional shortfall.

The uncertainty in the estimation of severity and cost effectiveness should be integrated to estimate the severity-adjusted probability of being cost effective.

1 Introduction

In the context of priority setting in health care, an important issue is whether or not to give more weight to health gains in particular circumstances or beneficiaries. Important examples include gains at the end of life, in the young or in severely ill patients. If distributional concerns exist and are to be included in (decisions based on) economic evaluations, one way of doing so is by using differential cost-effectiveness thresholds based on equity classes. In the Netherlands, for instance, severity-based equity classes define three thresholds: a threshold of €80,000 per QALY applies to the most severe conditions with poor prognoses such as aggressive cancers, and two lower thresholds are used for conditions with a relatively better outlook. For example, a new heart failure treatment had to be evaluated against a €50,000 per QALY threshold according to the guidelines [1]. The estimate of severity, however, is uncertain and this can cause issues in the evaluation and interpretation of cost-effectiveness results. For example, in April 2019 the appraisal committee of the Dutch National Health Care Institute had difficulties evaluating the cost

✉ Matthijs M. Versteegh
versteegh@imta.eur.nl

¹ Institute for Medical Technology Assessment, Erasmus University of Rotterdam, Burgemeester Oudlaan 50, 3000 DR Rotterdam, The Netherlands

² Erasmus School of Health Policy & Management, Erasmus University of Rotterdam, Burgemeester Oudlaan 50, 3000 DR Rotterdam, The Netherlands

effectiveness of venetoclax as a treatment for chronic lymphocytic leukaemia. The committee noted that the applicable threshold for this condition could be either €80,000 or €50,000 per QALY due to uncertainty in the estimate of severity. Since the treatment would only be cost effective with the higher threshold, the committee noted that the uncertainty of the threshold caused great uncertainty about the cost effectiveness of the treatment [2].

Severity of illness generally reflects the (average) amount of health lost in a population affected by some illness. Using differential cost-effectiveness thresholds based on severity rather than a fixed threshold is assumed to result in a more equitable distribution of resources, despite a potential sacrifice in total aggregate health in society. Indeed, it seems that “people are, on the whole, willing to sacrifice aggregate health in order to give priority to the severely ill” [3]. However, operationalizing severity into measurable units is not straightforward [4] and decision makers and scholars alike are in disagreement on the optimal approach. The disagreement entails both the principles underlying severity weighting (i.e. how to define who is worse off) and the operationalization of the principle (i.e. how to come from the principle to quantitative expressions of severity) [5].

Several governments and relevant authorities have instruments in place to set priorities based on principles other than health maximization. In the UK, incremental cost-effectiveness ratios above the upper threshold range of £30,000 may be acceptable for life-extending end-of-life treatments meeting certain criteria, such as extending life by at least 3 months [6]. The need to take severity into account is also reflected in health policies in countries such as Finland, France, Germany, Spain and Sweden [2], although not necessarily in a quantified manner or directly related to differential cost-effectiveness thresholds.

The Netherlands and Norway both explicitly integrate severity in their decision-making process but have proposed differential thresholds based on two distinct principles and operationalization of severity. In the Netherlands, a decision-making framework has been adopted with three primary criteria: necessity, effectiveness and efficiency (cost effectiveness). Equity considerations are integrated in this framework within the necessity criterion (the [medical] need to insure the intervention) and in the efficiency criterion through differential cost-effectiveness thresholds for increasing severity [5]. There are three increasing cost-effectiveness thresholds, which are associated with three increasing categories of ‘severity’ of the condition that the intervention under evaluation targets [7]. The severity categories are based on the proportion of quality-adjusted life-years (QALYs) lost relative to quality-adjusted life expectancy (QALE) of matched (on

age and gender) non-ill individuals, a principle referred to as *proportional shortfall* [8]. Recently, a review of the application of proportional shortfall in the Netherlands was published [9]. Aside from the Netherlands, Norway also envisions a future with a differential threshold based on severity [10], and has proposed a categorical and continuous severity criterion based on absolute QALYs lost relative to QALE of matched (on age and gender) non-ill individuals, a principle referred to as *absolute shortfall*.

An often overlooked issue, which to our knowledge has not previously been addressed in the literature on severity-based differential thresholds, is the uncertainty and heterogeneity in the estimate of severity. Quantifications of severity that include estimations of future QALYs, regardless of the specific unit of measurement for severity, are uncertain due to several reasons, among which is uncertainty in future health of patients.

This paper first suggests standardizing the method to estimate two measures of severity: absolute and proportional shortfall. Second, a method is presented on how to combine the uncertainty of severity with the uncertainty in a cost-effectiveness model. This is done to address the fundamental problem decision makers face: since neither the exact incremental cost-effectiveness ratio nor the exact threshold can be known, we may need to estimate the severity-adjusted probability of being cost effective as an alternative to, or alongside, cost-effectiveness acceptability curves (CEACs) [11]. Finally, this paper proposes a method for doing so and highlights its use in the Netherlands in a stylized example.

2 Background

A healthcare intervention can be seen as welfare improving if its additional benefits outweigh its additional costs. Cost-utility analysis (CUA), “the primary method for evaluating policy under a utilitarian ethic” [12], expresses benefits in terms of QALYs and an associated monetary value of these QALYs, giving the following decision criterion for incremental net monetary benefit for adopting a new healthcare intervention:

$$v\Delta Q - \Delta C > 0, \quad (1)$$

where v is the consumption value of health, or the societal willingness to pay (WTP) for a QALY (distinctively different from the ‘ k -threshold’, which reflects the marginal returns on health spending or health opportunity costs from a healthcare perspective¹), ΔQ is the difference in QALYs

¹ Note that in using a healthcare perspective with different health opportunity cost thresholds to reflect equity concerns, similar issues arise as discussed here. Here, we assume, for simplicity’s sake, optimal and flexible budgets.

between the current standard of care and the new medical intervention and ΔC is the difference in societal costs between the current standard of care and the new medical intervention. The decision framework of Eq. 1 gives equal weight to all QALYs gained, irrespective of gaining them in those with poor or good health (or any other relevant distinction between groups). The formula can be adjusted to take into account differential thresholds based on equity considerations such as severity [5]:

$$v_i \Delta Q_i - \Delta C > 0, \quad (2)$$

The subscript i represents the possibility to consider the context of QALY gains and to attach different values to QALY gains occurring in different ‘equity classes’. Hence, when QALYs are gained in that specific equity class, a different monetary value of a QALY may apply to account for different societal WTP for the worse off, aiming to achieve a better balance between equity and efficiency [13]. Note that this is equal to attaching equity weights to the QALY gains when these equity weights reflect the relative societal value of these gains [14]. While many different elements may define an equity class [15], this paper limits the examples to equity classes based on severity. Also, it is worth mentioning that empirical estimates of both k -thresholds and v -thresholds are surrounded with uncertainty. For example, the Dutch estimate for k of €41,000 has a 95% credible interval of 25,900–110,400 [16], which could be taken into account when representing decision uncertainty. However, it typically is not the case that thresholds defined in a country are a one-on-one representation of empirical estimates and hence may also represent some fixed value or value range set by an appropriate authority.

A simplification in the often used notation in Eq. 2 is that it does not take into account the heterogeneity of equity classes *within* the population in which QALYs are gained. Simply put, Eq. 2 assumes homogeneity in patients in terms of equity class or implicitly uses an average, while not all patients in which QALYs are gained are equally worse off. In treating specific infections, for instance, some patients may be severely ill while others receiving the treatment are only mildly ill. To account for this variation, Eq. 2 can be rewritten to:

$$\sum_{j=1}^N \sum_{i=1}^n (v_{ij} \Delta Q_{ij} - \Delta C_{ij}) > 0, \quad (3)$$

to allow every patient j to belong to any of the i equity classes. However, only in patient-level simulation models will it be feasible to identify the individual equity class since only in those models a patient-specific value for Q and C is available. Therefore, we present an approach here that takes heterogeneity as well as uncertainty of equity classes into account using input from the probabilistic sensitivity

analysis, and is hence easier to implement in Markov-type models. To do so, we first need to define severity and then describe how to consistently estimate it. Note that it is irrelevant for the calculation whether equity classes are defined as continuous classes (many values for each i) or categorically defined classes.

2.1 Severity-Based Equity Classes in the Netherlands

In the Netherlands, efforts have been made to formally include severity-based equity classes in guidelines for economic evaluations in terms of proportional shortfall (PS), which describes absolute shortfall (AS) as a proportion of QALE. We first describe AS, then PS.

The absolute loss in QALE is referred to as AS:

$$AS = Q_n - Q_d. \quad (4)$$

Here we define Q_n as the country-specific general population average QALE, adjusted for age and gender of the patients (i.e. the health that could have occurred had the patient not been ill) and Q_d is the average patient-population-specific QALE (i.e. QALE in the indication area of the medical intervention).

In the Netherlands, three equity classes have been defined based on the (average) proportion of normal life expectancy at a given age that is lost in a population due to the relevant disease, referred to as PS [8]. The ‘population’ refers to the target population of a given intervention as described in, for example, a reimbursement dossier. In a letter to the Minister of Health, the Dutch National Health Care Institute proposed that for a PS between 10 and 40% of remaining QALE, a WTP threshold of €20,000 applies [7]. Similarly, between 41 and 70%, and > 71%, the threshold is, respectively, €50,000 and €80,000 per QALY. PS is calculated as described in Eq. 4 and expresses the number of QALYs lost in a population as a proportion of the number of QALYs that would be incurred in the general population matched for age and gender:

$$PS = \frac{AS}{Q_n}. \quad (5)$$

The relationship between AS and PS is.

$$AS = PS \times Q_n \quad (6)$$

Equations 4 and 5 describe the calculation of AS and PS. For several years, there was no clear guidance on how to calculate the constituting elements Q_n and Q_d , but in 2018 both Dutch and Norwegian reimbursement agencies have documented its calculation [17, 18], and hence for AS and PS there are policy documents available. However, both guidelines do not deal with uncertainty in the

estimates, and are not aligned in the methods to be used to calculate Q_n , both of which we address here.

First, we describe a consistent and reproducible way of calculating Q_n and suggest use of undiscounted QALYs in the comparator of the health economic model for the estimation of Q_d . Note that with equating Q_d to the undiscounted comparator in economic evaluations, we effectively state that population-specific QALE in the indication area of the condition should be calculated taking into account the current standard of care. This is different from earlier formulations that calculated Q_d by subtracting “a patient’s remaining QALY expectancy without treatment” from Q_n [8]. Hence, while we look at QALY losses occurring given current standards of care, previous definitions used the situation without any treatment as comparator. This choice will be elaborated on further below. Then, we describe how we integrate uncertainty from the health economic model with uncertainty in the estimate of Q_n .

3 Definitions for Quality-Adjusted Life Expectancy (QALE)

Here, we first describe a standardized estimation of Q_n and Q_d for use in AS and PS calculations and then proceed to suggest how the severity-adjusted cost-effectiveness threshold can be calculated for differential thresholds.

3.1 Normal QALE: Q_n

Normal-period QALE reflects what the average population can expect to incur in terms of numbers of QALYs remaining at any given age and gender. Hence, QALE requires an estimation of life expectancy and quality of life by age and gender. We propose to consistently estimate life expectancy using life tables and source of utility values for QALY computation that is consistent with the estimates for utility used in economic evaluation.

Here we use data and the methods protocol of the Human Mortality Database with 2013 country- and gender-specific mortality rates and a half-cycle correction [19]. The life tables are closed at 99+ years using a constant mortality risk from that age onwards. Quality of life is integrated in the life tables using EQ-5D-3L-based utility values stratified for 14 age groups (from 20 to 85 years) and gender in the life tables [20]. Quality adjustment is based on EQ-5D since, in the Netherlands, this tool is requested as the reference case for the computation of QALYs in economic evaluations. Thus, it is consistent to also use EQ-5D for the computation of Q_n , especially if Q_d is taken from the undiscounted number of EQ-5D-based QALYs for the comparator in the health

economic model. In estimation of QALE for this study, assumptions had to be made about quality of life for age groups < 20 years and > 85 years due to missing data. Quality of life for ages 0–20 years is assumed at 0.958, which is 0.02 higher than the Dutch EQ-5D-3L population norm for > 20- to 24-year olds and corresponds to the Dutch EQ-5D-5L estimate for the population aged 18–20 years [21]. For age > 85 years, the utility value is assumed to be the same as in the age group 80–85 years. We are aware that equivalence between EQ-5D-3L and 5L utility values is not established [22], but population norms for EQ-5D-5L are not (yet) available for the Netherlands.

Note that there are two elements of uncertainty in the estimation of Q_n : the applicability of the 2013 mortality rates to the population at the time of estimation and the heterogeneity in quality-of-life estimates. Since uncertainty regarding mortality rates is negligible, only heterogeneity regarding quality of life is included, detailed later on.

3.2 Patient-Population-Specific QALE: Q_d

Patient-population-specific QALE should be calculated for the population for which a cost-effectiveness model is submitted since this model reflects the population eligible for treatment. Hence, Q_d is equal to the modelled total number of undiscounted QALYs with standard of care at the mean age of the model population with country-specific quality-of-life values, which best represent patient QALE, using a lifetime horizon. Therefore, ‘standard of care’ is chosen as reference point rather than ‘no treatment’. Given that this is a deviation from traditional calculations, we highlight a few reasons why this was chosen. First, this estimate realistically reflects current survival and quality of life of those with the condition and hence provides the relevant background situation for the current decision to fund a new intervention. The better current treatment is, the lower the necessity to improve it even further. This is intuitive. Second, this proposed estimate aligns with the perspective of the decision maker who has to decide if the marginal benefit of the new treatment weighs up to its marginal costs, relative to the current treatment. Third, in the Netherlands, severity is calculated in the context of necessity of treatment, which refers to the claim on solidarity new treatments can make. It appears that, even though previous documents have referred to Q_d as QALE ‘with the condition’ in a more general sense [23], this interpretation of Q_d as ‘with the condition and current standard of care’ relates more closely to *additional* claims on solidarity. Fourth, given that equity weights often aim to (somehow) reflect societal preferences for distribution, it seems that a realistic alternative (like the comparator) would better align with preferences for treating particular diseases than hypothetical ones. Finally, finding estimates of quality

of life in case of not treating a disease may be difficult for all diseases for which good treatments exist.

As a consequence, the value of Q_d and the severity of disease in the relevant population is dynamic and follows developments in current standard of care. When direct access to the economic model is available, the estimates of the probabilistic sensitivity analysis can be used to reflect the distribution (capturing both heterogeneity and uncertainty) of potential estimates for Q_d .²

To provide an example illustrating the above and the impact of uncertainty, consider a new therapy (T) that was developed for a condition in a population of males with a mean age of 33 years, for which current standard of care results in an average QALE (Q_d) of 12.25 years. Life tables combined with Dutch quality-of-life data suggest that normal QALE (Q_n) for 33-year-old males is 42.83 years. Following Eqs. 4 and 5, the point estimate for PS is 0.71 ($42.83 - 12.25/42.83$) and the point estimate for AS is 30.58. Following the Dutch guidelines, a cost-effectiveness threshold of €80,000 would apply.

However, estimates of PS and AS are inherently uncertain, since they are calculated using both Q_n and Q_d , which are both uncertain. When a point estimate for PS is used to determine the applicable WTP threshold, small differences can have a sizable impact. In the context of treatment T, a PS of 71% (e.g. $Q_d = 12.25$ and $Q_n = 42.83$) implies a Dutch WTP threshold of €80,000 to be applicable. However, a value of Q_d that is just slightly larger (e.g. $Q_d = 13$, $PS = 42.83 - 13/42.83 = 0.7$, $AS = 42.83 - 13 = 29.83$) results in a WTP threshold of €50,000 in the Netherlands, while the two point estimates of Q_d in this example may lie well within the confidence interval and may even be equally likely. Similarly, the value of Q_n is also uncertain, and may impact the relevant threshold to be applied and is better represented by a distribution than by a point estimate.

To address the uncertainty in Q_n , the country-, age- and gender-specific mortality rates are combined with utility values from a random draw of k replications (here 1000) from a beta distribution using country-specific published EQ-5D-3L standard errors by age and gender [24], resulting in 1000 estimates of Q_n . The uncertainty for Q_d is included through the probabilistic sensitivity analysis of the health economic model or can alternatively be sampled using published mean

and standard deviation of the undiscounted QALYs in the comparator of the model. A simple disease burden calculation tool (the iMTA Disease Burden Calculator, iDBC) to perform these calculations by giving age, gender and the number of undiscounted QALYs from the comparator is available from the authors.

4 Severity-Adjusted Probability of Being Cost Effective

The severity-adjusted probability of being cost effective can be calculated by combining the results of the sampled estimates of Q_n and Q_d with the results of the probabilistic sensitivity of a cost-effectiveness model to calculate the net monetary benefit at given values for PS and the associated threshold using the equations below. That resulting value represents the severity-adjusted probability of being cost effective (SAPCE).

$$SAPCE_s = \frac{1}{n} \sum_{s=1}^n Y_s, \quad (7)$$

where

$$Y_s = \begin{cases} 1 & \text{if } v_s \times \Delta Q_s - \Delta C_s \geq 0 \\ 0 & \text{if } v_s \times \Delta Q_s - \Delta C_s < 0 \end{cases}.$$

Table 1 illustrates the application of SAPCE using a selection of hypothetical outputs of a health economic model. In the table, we use the Dutch proposal of PS as an indicator of severity, where v_s is the PS-based applicable threshold for the sampling estimate s , which can take on the value €20,000, €50,000 or €80,000 depending on the value of PS. Note that the method would also apply to continuous differential thresholds, in which case the applicable threshold v in Table 1 would be a unique value for each estimate of severity rather than a categorical variable with only three possible values. Note also that when the PSA results in a skewed distribution, the deterministic estimate of severity may well be different from the probabilistic estimate of severity.

4.1 Example Application

The implications of using (and not using) the SAPCE procedure in jurisdictions that wish to apply equity weighting is demonstrated here. To do so, we introduce a stylized example for a hypothetical treatment in, say, oncology in the Netherlands and evaluate it using differential thresholds based on the current proposed methods in those countries and using the new SAPCE to highlight relevant differences to the decision-making context.

² Note that normative choices regarding the appropriate population to calculate severity in as well as the relevant time frame for calculating it, remains important. In vaccination, for example, health effects are far in the future, uncertain, and affect only a few of the total treated population. Hence, calculating severity for preventive treatments requires the determination of the appropriate population after which the methods presented in this paper can be applied. A discussion that recognizes the complexity of the issues relating to these normative choices is outside the scope of this paper.

Table 1 Example calculation of the severity-weighted cost effectiveness

Model run (s)	Disease burden calculation				Model results			Combined results		
	QALE patients (Q_d)	QALE general population with same age and gender (Q_n)	Absolute shortfall ($Q_n - Q_d$)	Proportional shortfall $[(Q_n - Q_d)/Q_n]$	Applicable threshold (V_s)	Incremental costs, € (ΔC)	Incremental benefits (ΔQ)	ICER, € ($\Delta C/\Delta Q$)	INMB, € ($\Delta Q \times V_s - \Delta C$)	Cost-effective? ^a
1	15	25	10	0.40	20,000	20,000	0.60	33,333	-8000.00	0
2	16	24	8	0.33	20,000	8000	0.50	16,000	2000.00	1
3	17	28	11	0.39	20,000	15,000	0.60	25,000	-3000.00	0
4	15	28	13	0.46	50,000	10,000	0.50	20,000	15,000.00	1
5	14	27	13	0.48	50,000	10,000	0.40	25,000	10,000.00	1
6	13	26	13	0.50	50,000	25,000	0.30	83,333	-10,000.00	0
7	15	26	11	0.42	20,000	25,000	0.60	41,667	-13,000.00	0
8	15	32	17	0.53	50,000	15,000	0.50	30,000	10,000.00	1
9	16	25	9	0.36	20,000	25,000	0.60	41,667	-13,000.00	0
10	16	26	10	0.38	20,000	20,000	0.80	25,000	-4000.00	0
Severity weighted probability of being cost effective										40%

ICER incremental cost-effectiveness ratio, INMB incremental net monetary benefit, QALE quality-adjusted life expectancy

^a 1 = Yes

We use a hypothetical Markov model for an oncological treatment at old age (88 years, 50% males) with three health states (progression, progression free and dead). The model compares current care with a new hypothetical cancer treatment that delays time to progression by about 50%, with monthly costs of €110 before progression and €350 after progression, and a utility increase of 0.06 in the progression-free period for the treated group. Probabilistic sensitivity analyses are conducted using assumptions on the standard errors distributions for the input parameters.

The cost-effectiveness results of the hypothetical oncology model with population age of 88 years and 50% males are summarized in Table 2.

Estimates of Q_n are provided for the Netherlands using country-specific mortality rates and quality-of-life estimates corrected for age and gender with confidence intervals analytically derived from the 1000 samplings as described in the sections above. At age 88 years, 50% males, Q_n is 3.84 [95% CI 3.64–4.02]. The estimate of the mean Q_d is obtained from the probabilistic sensitivity analysis conducted in the hypothetical cost-effectiveness model. Standard of care results in a value for Q_d of 1.0976 QALYs. From these values, the mean AS and PS were calculated including 95% confidence intervals, the results of which are provided in Table 3.

According to a deterministic analysis for the Dutch setting, the appropriate WTP threshold for the hypothetical new oncology treatment would be €80,000 since $PS = 0.71$ ($2.74/3.84 = 0.71$). However, Table 3 also shows that the confidence interval for the point estimate of PS includes values < 0.71 , which are associated with a €50,000 threshold. Thus, the applicable threshold might also be €50,000 (the €20,000 threshold was not relevant with a probability of zero). In this case, we observed that 340 out of 1000 sampled values of PS were 0.4–0.7 and 660 were ≥ 0.71 .

Assuming a threshold/WTP of €80,000 per QALY, the probability of the new intervention being cost effective is about 85%. Assuming a threshold/WTP of €50,000 per QALY, the probability is about 3%. The severity-adjusted probability of being cost effective is 62.4%. Figure 1 shows the standard CEACs for the standard of care (dotted decreasing line) and the intervention (solid increasing line) with the two possible thresholds and $SAPCE = 0.624$ (probability of being cost effective given all probable thresholds of 62.4% according to the methods described in Table 1). Clearly, the use of SAPCE reduces the chance of an incorrect coverage decision due to the technology being evaluated against the ‘wrong’ threshold. Note that the value for SAPCE is placed on the Y axis of Fig. 1 to illustrate the difference with the probability of being cost-effective with traditional CEACs and a deterministic estimation of PS.

Table 2 Hypothetical cost-effectiveness model 0% discounting

	Incremental costs (€)	Incremental QALYs
New treatment	58,432.00	1.8572
Standard of care	10,445.00	1.0976
Incremental	47,987.00	0.7596
ICER	€63,176.53	

ICER incremental cost-effectiveness ratio, QALYs quality-adjusted life-years

Table 3 Loss in quality-adjusted life at age 88 years with 1.0976 QALYs remaining

	Absolute shortfall [95% CI]	Proportional shortfall [95% CI]
Netherlands	2.74 [2.67–2.82]	0.71 [0.69–0.74]

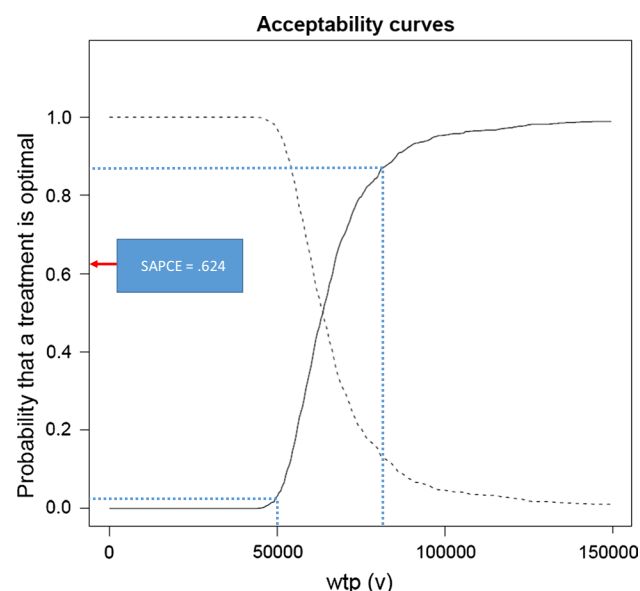


Fig. 1 Cost-effectiveness acceptability curves at willingness to pay (WTP)=50,000 and 80,000 and severity-adjusted probability of being cost effective (SAPCE) of the hypothetical oncology treatment

5 Discussion

This paper argues that countries and decision makers that wish to apply differential categorical thresholds, based on some equity criterion like severity of illness, need to account for heterogeneity and uncertainty in the point estimates of the appropriate measure, in relation to the appropriate threshold. A severity-adjusted probability of being cost effective may better reflect the uncertainty and heterogeneity regarding the calculations than a point estimate does.

Reflecting this may lead to better informed decisions that take into account both efficiency and equity considerations.

The approach presented here integrates the uncertainty in the estimation of appropriately expressed severity of illness and the uncertainty in the incremental cost-effectiveness ratio. We presented a standardized approach to address uncertainty in two common ways of expressing severity of illness—AS and PS. We showed how this uncertainty can be integrated with probabilistic sensitivity analyses.

The hypothetical cost-effectiveness model showed the relevance of appropriately expressing uncertainty for these parameters. In the Dutch situation, a minor change in life expectancy with the disease (i.e. 2.70 QALYs lost instead of 2.74) slightly lowered the PS estimate (within its confidence interval), causing it to fall into another threshold category—thus substantially lowering the applicable cost-effectiveness threshold. The associated probability of being cost effective could range from 85 to 3% for estimates of PS that were all within the same confidence interval. We argue that neither estimate would be correct and suggest that the severity-adjusted probability of being cost effective, which was 62.4% in the example, is a better reflection of the probability of making a correct decision regarding the adoption of the technology. Arguably, the use of categorical thresholds rather than continuous ones increases the importance of integrating uncertainty since the categorization in a specific equity class can have a large impact on the probability of being cost effective.

There are several limitations to our study. The first is the selection of severity measures used to exemplify the procedure. PS is the method of choice in the Dutch guidelines as an operationalization of severity in the context of reimbursement decisions. Different operationalizations have some support in the Netherlands, Norway and the UK, and all incorporate both a time and quality component [25]. The use of the PS method here follows the Dutch guidelines and does not express a preference of this method over another by the authors.

The second limitation is difficult to correct, and arguably a small flaw in the calculation of AS and PS as described here. Age- and gender-specific QALE for the non-ill (Q_n) is calculated based on average mortality and quality-of-life data that includes the population for which severity is being calculated (Q_d). A truly unbiased estimate of Q_n should be based on mortality and quality-of-life data that excludes the population for which Q_d is calculated. Note that this limitation is not unique to our proposed method, but present in any calculation of AS and PS based on population norms. Moreover, unless the population under evaluation accounts for a notable part of mortality and morbidity in the general population, the resulting bias is arguably negligible. Should decisions need to be made for treatments targeting a highly common condition, such

that the estimate of Q_n would be noticeably biased, corrections could be made to exclude this bias.

A further limitation is that we used mortality figures from 2013; at the time of analysis these were the most recent figures that were available for each of the countries included in this study in the Human Mortality Database. Also, the age- and gender-specific quality-of-life values were taken from the EQ-5D-3L, which has since been replaced with EQ-5D-5L [26]. Such practical limitations can be solved by updating figures as soon as more recent or better ones become available.

The incorporation of a severity-based cost-effectiveness threshold has limitations in and of itself that are not addressed by taking uncertainty into account. We have worked under the simplifying assumption that we use an optimal, flexible budget and social WTP as the appropriate threshold. In case of fixed, non-optimal budgets, displacement of existing technologies and their health opportunity costs play a role in the decision [27]. In such a context, either in a narrow healthcare perspective or a broader societal perspective, the use of equity weights or differentiated valuations of health gains are relevant for both the new healthcare intervention as well as the displaced activities. More often than not, governments are unable to identify the displaced healthcare interventions and, hence, their equity weight or social value; this introduces a practical issue [28]. As with estimating average opportunity costs of marginal spending, the best estimate may be the average equity weight of current spending. This is an important avenue for future research.

Our study also has some strengths that are worth mentioning and overcomes the limitations of a recent publication regarding the weighting of QALY gains to account for distributional concerns [29]. Our study proposed a way to address uncertainty and heterogeneity in QALE estimates. These methods are robust and do not require strong assumptions about Q_n . Also, the use of EQ-5D as a source for quality-of-life values aligns the QALY calculations for severity with those of the health economic model. Finally, our study keeps the categorical nature of three increasing cost-effectiveness thresholds intact, which facilitates communication and avoids linear interpolations between and extrapolations beyond threshold values. The methods produced here can, however, also be applied to continuous estimates of severity rather than categorical ones.

The methods presented here are a quantitative approach to the description of severity in the context of reimbursement decisions. In order to enable the quantification, the severity criterion has to be well defined and measurable. Note that this is true for all quantifications, also when used less sophisticatedly.

Our approach can be used in the assessment phase and as such does not preclude the use of additional qualitative

and/or deliberative assessments of severity in an ‘appraisal-like’ setting. Rather than integrating uncertainty surrounding the severity estimate with uncertainty from the health economic model, decision makers may prefer to keep these factors separate to facilitate a deliberative ‘weighing’ of economic and equity considerations. However, also in that situation a proper estimate of severity including its uncertainty is required. Our approach simply acknowledges that when a quantified approach to incorporating severity is adopted, it is important to take into account uncertainty around the parameters that make up the calculation. The severity-adjusted probability of being cost effective as presented here is suitable for that task.

6 Conclusion

Differential thresholds can represent differential societal WTP for QALY gains in those who are more severely ill. Estimates of severity, regardless of the underlying severity principle, are uncertain, and this uncertainty needs to be integrated with the uncertainty of the incremental cost-effectiveness ratio. Here, we presented a procedure to consistently estimate severity across reimbursement dossiers, and a procedure to integrate the uncertainty of that estimate with the uncertainty in the incremental cost-effectiveness ratio. The resulting severity-adjusted probability of being cost effective is better capable of informing decision makers than one that does not take the uncertainty of severity estimates into account.

Author contributions MV, IR, NCB, AA, VRD and WB identified relevant analyses, drafted and critically reviewed the manuscript. MV conceived the research plan and conducted the analyses.

Compliance with Ethical Standards

Funding This paper was not funded.

Conflict of interest The authors MV, IR, NCB, AA, VRD and WB do not have a conflict of interest to report.

Data availability An open access R tool to calculate absolute and proportional shortfall according to the methods presented in this paper can be found here: <https://imta.shinyapps.io/iDBC/>.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ramos IC, Versteegh MM, de Boer RA, Koenders JM, Linssen GC, Meeder JG, Rutten-van Mölken MP. Cost effectiveness of the angiotensin receptor neprilysin inhibitor sacubitril/valsartan for patients with chronic heart failure and reduced ejection fraction in the Netherlands: a country adaptation analysis under the former and current Dutch Pharmacoeconomic Guidelines. *Value Health*. 2017;20(10):1260–9. <https://doi.org/10.1016/j.jval.2017.05.013>.
- ACP public meeting 05-04-2019. <https://www.zorginstituutnederland.nl/binaries/zinl/documenten/adviezen/2019/04/05/acp-advies-over-venetoclax/ACP-advies+over+venetoclax.pdf>. Accessed 23 Apr 2019.
- Shah KK. Severity of illness and priority setting in health-care: a review of the literature. *Health Policy*. 2009. <https://doi.org/10.1016/j.healthpol.2009.08.005>.
- Wailoo A, Tsuchiya A, McCabe C. Weighting must wait. *Pharmacoeconomics*. 2009. <https://doi.org/10.2165/11314100-00000-00000-00000>.
- van de Wetering EJ, Stolk EA, van Exel NJ, Brouwer WBF. Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall. *Eur J Health Econ*. 2013;1:107–15. <https://doi.org/10.1007/s10198-011-0346-7>.
- NICE. Appraising life-extending, end of life. 2009. National Institute for Health and Clinical Excellence. <https://www.nice.org.uk/guidance/gid-tag387/documents/appraising-life-extending-end-of-life-treatments-paper2>. Accessed 11 May 2019.
- Zwaap J, Knies S, Van der Meijden C, Staal P, van der Heiden L. Cost-effectiveness in practice. 2015. <https://English.zorginstituutnederland.nl/publications/reports/2015/06/16/cost-effectiveness-in-practice>. Accessed 18 Jan 2017.
- Stolk EA, Van Donselaar G, Brouwer WBF, Busschbach JJV. Reconciliation of economic concerns and health policy: illustration of an equity adjustment procedure using proportional shortfall. *Pharmacoeconomics*. 2004. <https://doi.org/10.2165/00019053-200422170-00001>.
- Reckers-Droog VT, van Exel NJA, Brouwer WBF. Looking back and moving forward: on the application of proportional shortfall in healthcare priority setting in The Netherlands. *Health Policy*. 2018. <https://doi.org/10.1016/j.healthpol.2018.04.001>.
- Ottersen T, Førde R, Kakad M, Kjelleveid A, Melberg H, Moen A, Ringard A, Norheim O. A new proposal for priority setting in Norway: open and fair. *Health Policy*. 2016. <https://doi.org/10.1016/j.healthpol.2016.01.012>.
- van Hout BA, Al MJ, Rutten FFH. Costs, effects and *cle*-ratios alongside a clinical trial. *Health Econ*. 1994. <https://doi.org/10.1002/hec.4730030505>.
- Ruger JP. Health and social justice. Oxford: Oxford University Press; 2009.
- Wagstaff A. QALYs and the equity-efficiency trade-off. *J Health Econ*. 1991. [https://doi.org/10.1016/0167-6296\(91\)90015-F](https://doi.org/10.1016/0167-6296(91)90015-F).
- Bobinac A, van Exel NJ, Rutten FFJ, Brouwer WBF. Inquiry into the relationship between equity weights and the value of the QALY. *Value Health*. 2012. <https://doi.org/10.1016/j.jval.2012.07.002>.
- Norheim OF, Baltussen R, Johri M, et al. Guidance on priority setting in health care (GPS-health): the inclusion of equity criteria not captured by cost-effectiveness analysis. *Cost-effectiveness and resource allocation*. 2014. <https://doi.org/10.1186/1478-7547-12-18>.
- Van Baal PHM, Perry-Duxbury M, Bakx P, Versteegh MM, van Doorslaer E, Brouwer WBF. A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. *Health Econ*. 2018. <https://doi.org/10.1002/hec.3831>.
- Legemiddelverket Guidelines for the documentation for single technology assessment (STA) of pharmaceuticals. Norway: s.n., 2018. <https://legemiddelverket.no/Documents/English/Public%20funding%20and%20pricing/Documentation%20for%20STA/Guidelines%20151018.pdf>. Accessed 11 May 2019.
- Zorginstituut. Ziektelast in de praktijk. De theorie en praktijk voor het berekenen van ziektelast bij pakketbeoordelingen. The Netherlands: s.n., 2018. https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2018/05/07/ziektelast-in-de-praktijk/Ziektelast+in+de+praktijk_definitief.pdf. Accessed 11 May 2019.
- Wilmoth JR, Andreev K, Jdanov D, et al. Methods protocol for the human mortality database. University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock, pp. 10–11. 2007. <http://mortality.org>. Accessed 31 May 2007.
- Heijink R, van Baal PHM, Oppe M, Koolman X, Westert G. Decomposing cross-country differences in quality adjusted life expectancy: The impact of value sets. *Popul Health Metr*. 2011. <https://doi.org/10.1186/1478-7954-9-17>.
- Versteegh MM, Vermeulen KM, Evers SMA, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five level version of EQ-5D. *Value Health*. 2016. <https://doi.org/10.1016/j.jval.2016.01.003>.
- Hernández-Alava M, Pudney S. Econometric modelling of multiple self-reports of health states: the switch from EQ-5D-3L to EQ-5D-5L in evaluating drug therapies for rheumatoid arthritis. *J Health Econ*. 2017;1:139–52. <https://doi.org/10.1016/j.jhealeco.2017.06.013>.
- Couwenbergh B, van de Meer F, Weghaus-Reus S, Schelleman H, Zwaap J. Pakketbeheer in de praktijk deel 3. s.l.: Zorginstituut Nederland. 2013. <https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2013/10/18/pakketbeheer-in-de-praktijk-deel-3/Pakketbeheer+in+de+Praktijk+%28deel+3%29.pdf>. Accessed 11 May 2019.
- Cabases J, Janssen B, Szende A. Self-reported population health: an international perspective based on EQ-5D. Rotterdam: Springer; 2014.
- Nord E, Johansen R. Concerns for severity in priority setting in health care: a review of trade-off data in preference studies and implications for societal willingness to pay for a QALY. *Health Policy*. 2014. <https://doi.org/10.1016/j.healthpol.2014.02.009>.
- Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011. <https://doi.org/10.1007/s11136-011-9903-x>.
- Claxton K, Paulden M, Gravelle H, Brouwer WBF, Culyer AJ. Discounting and decision making in the economic evaluation of health care technologies. *Health Econ*. 2011. <https://doi.org/10.1002/hec.1612>.
- Paulden M, O'Mahony JF, Culyer AJ, McCabe C. Some inconsistencies in NICE's consideration of social values. *Pharmacoeconomics*. 2014. <https://doi.org/10.1007/s40273-014-0204-4>.
- Hernæs UJ, Johanson KA, Ottersen T, Norheim OF. Distribution-weighted cost-effectiveness analysis using lifetime health loss. *Pharmacoeconomics*. 2017;35:965. <https://doi.org/10.1007/s40273-017-0524-2>.