Hacienda Pública Española/Review of Public Economics 242-(3/2022): 111-127
© 2022, Instituto de Estudios Fiscales
https://dx.doi.org/10.7866/HPE-RPE.22.3.5

QALY Maximization and the Social Optimum

GUILLEM LÓPEZ CASASNOVAS*
Universitat Pompeu Fabra

JOSÉ LUIS PINTO PRADES**
Universidad de Navarra

Received: July, 2021
Accepted: February, 2022

Abstract

The quality-adjusted life year (QALY) approach as a paradigmatic measure of health outcomes is widely used in the economic evaluation of health policies and is even trying to be emulated in other public policy areas. The objective of this paper is to structure the arguments underlying the use of QALYs in cost-effectiveness analyses, which condition its applicability. To this effect, the paper intervenes in the unreasonable use of the approach in terms of its limitations when prioritizing health services. In short, the authors are generally in favor of the instrument but do not have blanket enthusiasm in support of it.

Keywords: QALY, Cost effectiveness, Thresholds, Economic evaluation, Pharmaeconomics.

JEL Classification: H51, I31, I18, K23, I14.

Introduction

The quality-adjusted life year (QALY) is a measure of health status that considers both quantity and quality of life. One QALY is the equivalent of one year in a perfect state of health. If an individual's health is below this maximum level it is adjusted with respect to death, which is assigned the value 0. In some cases, it is even possible to account for negative QALYs to reflect states of health that are considered "worse than death". This approach assumes that health is dependent upon both the length and the quality of life, combining the values assigned to the two in a single numerical index. These values are derived from collective preferences regarding certain states in terms such as pain, anxiety, mobility, and other psychometric factors. The value associated with a given state of health is multiplied by the number of years of life in that state of health to determine the total number of QALYs.

* ORCID ID: 0000-0001-8020-9987.

** ORCID ID: 0000-0002-9684-3410.

The quality-adjusted life year (QALY) has become consolidated as a paradigmatic measure to estimate health outcomes in the economic evaluation of health policies in terms of their cost effectiveness. Nowadays, it is widely applied, albeit without some controversy. We believe that one of the main reasons of its popularity is that it is a very simple (apparently) way of solving difficult problems of resource allocation. Once groups such as the Euroqol produce a national set of values ("tariffs" as they are known in the health economics literature) of health states, it is not difficult to calculate QALYs without being an expert on the QALY model. As we will show in this paper, the simplicity of the model does not come without problems. Since the model is so simple it cannot be a good descriptive measure of many health treatments. Especially when (as it is usually the case) health problems are not chronic. However, it is very hard, even for people who know the literature, to tell when the limitations of the model make the QALY a good enough or a very bad approximation to the real value of health. We hope to clarify this point during the rest of the paper.

Health economists' main contribution nowadays largely revolves around the cost-effectiveness thresholds[1] of treatments (and not just their actual effectiveness), which are then applied to the adoption of health policies or the price and reimbursement of medications. However, the method used for their calculation has some caveats, indicating that its results must be used with caution and never be linked to rules of thumb[2] in paramount societal matters such as whether or not to fund certain health services or to exclude some of their beneficiaries. Furthermore, its indiscriminate application can lead to the adoption of socially unacceptable decision-making procedures in specific cases (regarding treatments, beneficiaries, and guidelines) as the basis of a bad management of public interest.

NICE *vs.* IPT: What is more rational?

We will start this overview by taking the drug Eluxadoline as an example. The National Institute for Health and Care Excellence (NICE), the English health evaluation agency, recommended its use for irritable bowel syndrome (National Institute for Health and Care Excellence, 2017) because its cost per QALY was around £12,000, which is way below the £30,000 threshold. However, the Spanish Therapeutic Positioning Report (TPR) (Informe de Posicionamiento Terapéutico-IPT) (Agencia Española de Medicamentos y Productos Sanitarios, 2019), compiled by the General Directorate responsible for the country's Ministry of Health services, concluded that this medication should not be funded. The NICE based its decision on the cost-effectiveness model, reaching the conclusion that "the ICER[3] was likely to be within the range normally considered a cost-effective use of NHS resources". The question is, can we conclude that the NICE decision is right, given that the cost per QALY is way below the threshold, and that the TPR is wrong? The aim of this paper is to explain why this conclusion is not necessarily correct and what the implications are for applying the cost per QALY to economic evaluation in Spain.

Economic evaluation is an example of what Sugden, 2021 called "constructivist rationality". Constructivist rationality assumes that there is a single rational decision, a unique "optimum" in making individual and social decisions. The neoclassical model assumes that when individuals allocate their private resources in the market they do so in an "optimum" way, given that they are able to balance the marginal benefit cost ratio ($\frac{MU_X}{P_X} = \frac{MU_Y}{P_Y}$) in all

their consumer decisions. This conclusion stems from the assumption that consumers have all the available information about their preferences and their budgetary constraints (income and prices). If we have the pertinent information, then the problem of allocating resources can be easily resolved using constrained optimization methods.

Regarding public resources, there are often neither prices nor markets. Given the lack of an explicit market, economists try to obtain information to achieve this optimum using other methods. They need to ascertain the preferences of individuals and this is achieved either with revealed preferences methods[4] or stated preferences methods[5]. In the case of the economic evaluation of health, the main method is the latter. The reason is that in the absence of a health care market in most countries it is not possible to observe the implied monetary value of health. Once the preferences of individuals are known, we can calculate the marginal benefit cost ratios (or cost per QALY) and "reach the highest social wellbeing as the sum of the wellbeing of each individual, maximizing the sum of quality-adjusted life years (QALY) in the population" (Campillo-Artero, C., 2013).

The way in which this approach has become a mainstay in health economics has a specific name: The cost per QALY threshold. We identify preferences using QALYs and social optimum with its maximization, thus showing clearly how to answer the question in the way NICE did. Since the social optimum is achieved using the cost per QALY threshold, it is rational to finance the medication if the threshold is £30,000 and the cost per QALY is £12,000. The decision not to fund the medicine would be irrational. Nonetheless, the assumptions that underpin this assertion are extremely debatable, as we will now explain.

1. QALYs as a reflection of individual preferences

QALYs are an extremely simplified model of individual preferences. In fact, we are nowadays still using the model developed by Fanshel (Fanshel, 1972; Fanhsel and Bush, 1970 and Bush *et al.*, 1972) at the beginning of the 1970s. In these publications, the QALY model was presented in the same way as it is used currently: each state of health is assigned a constant value (between 0 and 1) and is weighted by the duration. Total health is the sum of the years weighted by this value. This is known today as the "linear QALY model". Torrance *et al.*, 1972 used a scheme similar to that of Fanshel and Bush, although with new methods (Time Trade-off and Standard Gamble) more firmly grounded in economic theory to calculate the weightings by quality of life. This model was proposed intuitively, grounded in the authors' assumption that it was an effective way to model health; it was not based on the observation that it reflects preferences. In fact, in these pioneering studies the authors are well aware of the limitations (e.g., linearity in life years or in the aggregation of QALYs) of the model (Fanshel and Bush[6], 1970; Torrance, 1973). QALYs stem from the imposition of a specific function and not from observation of the population's preferences. Pliskin *et al.* (1980) formalized QALY's as functions of individual decision making for the case of chronic states. They showed that the linear QALY model assumes mutual utility independence between quality of life and life years[7], constant proportional trade-offs[8] and risk neutrality[9].

The point that we are making is that researchers proposed the linear QALY model because it was simple and intuitively appealing. They did not started studying individual preferences and proposing the model that better reflected those preferences. Regarding health problems that vary with the passage of time, which is the most usual case, the main assumption is what Wakker (2008) calls "independence of disjoint health states". In other words, the use of a state of health depends on neither the past nor its future evolution. This assumption is highly implausible. In short, fifty years later and after many academic publications (Abellán *et al.*, 2016), the QALY model used nowadays is the same one as was proposed at the beginning of the 1970s.

The question we ask is why we are still using such a restrictive model. In our opinion, its main attraction is its simplicity, explaining its widespread use. For example, a treatment for metastasized cancer increases life expectancy by some months and improves some aspects of quality of life, but it produces different types of side effects. Aggregating all these effects can be extremely complex. However, the algorithm used to convert this problem into a QALY is relatively simple. Everything is linear and additive, and so QALYs have become a simple way to provide an apparently clear response to a complex question. However, having said this, in no way do clarity and simplicity guarantee that this is the right solution or, even less, the "optimum" one. The secret to resolving a complex problem using a simple algorithm is that simplicity is achieved through a series of assumptions. Once we apply the algorithm and we obtain "the number" it is very difficult to know the point to which the model's assumptions influence the final result.

2. Information about individual preferences

The linear QALY model, which is currently the most applied one, needs just one parameter for its use: the utilities of health states. There appears to be a certain consensus that these must reflect society's preferences, which is why surveys are carried out with groups representative of the general population. The problem that researchers have found is that these values are very manipulable. It is true that some methods seem to produce more consistent results than others. For example, we have shown that non-transparent methods seem to elicit preferences that are closer to Expected Utility than transparent methods (Pinto Prades *et al.*, 2018). In non-transparent methods indifference is reached through sequences of choices that hide the subject the goal of the sequence, i.e., to find indifference between two health profiles. This (apparently) reduces anchoring bias and improves the consistency of the valuations. However, these surveys ask extremely difficult questions, and it is easy to introduce apparently irrelevant changes in the design of a survey that can produce extremely different results. This matter has been examined in detail in Pinto Prades *et al.* (2019). In brief, the problems are:

1. Order effects: if we determine the utility of state A first and then the utility of state B, we obtain a different utility from if we do so in the opposite order (Pinto-Prades *et al.*, 2019b).

2. Starting point effects: the use of the state of health depends on the first task the subjects must do to reach the point of indifference required by the different methods (Augestad *et al,*, 2016).

3. Inversion of preferences: depending on the way the question is formulated, the utility of state of health A can be greater or less than that of state B (Bleichrodt and Pinto-Prades, 2009).

4. Direct or indirect valuation of gains in health: improvement in health from state B to state A is usually measured like the rest of the utilities, using U(A)-U(B). On the other hand, when we find this difference in a direct way (Taylor *et al.*, 2017) we obtain a different result.

5. Internal inconsistencies within the methods: two theoretically equivalent ways of asking produce different results. Therefore, it is easy to obtain two utilities for the same subject (Bleichrodt *et al.*, 2003).

6. Dependency on the health profile: the utilities change with the health profile used to describe the state of health. For example, Euroqol and the Health Utility Index (HUI) produce different relative uses, and so the incremental cost, the QALY cost, depends on the health profile used (Hanmer *et al.*, 2016).

7. The utilities for the same health problem vary enormously (Zhou *et al.*, 2021).

Many of these problems are not observed in many surveys used to obtain utilities, simply because the study is not designed to observe these effects. A paradigmatic case was the Euroqol 5D-5L in the UK. Given the problems observed by Hernández Álava *et al.* (2018) in Euroqol 5D-5L for the UK, NICE recommended not using the utilities obtained by Devlin *et al.* (2018). However, and this is our main point, this decision was not easy to make because the inconsistencies observed were not reflected in any immediately obvious way in the regression presented by the authors. NICE had to consult with a group of six external specialists to evaluate the results and reach the conclusion to not recommend the tariff produced by Devlin *et al.* (op. cit.), despite the evident limitations of the data.

In conclusion, the ease with which the answers to questions to evaluate health states can change suggests that the members of the general population have very vague and imprecise preferences about these values. Therefore, one of the assumptions (well-defined preferences) to calculate the "optimum" is extremely problematic. This is not surprising, given that rather complicated scenarios are presented in these surveys with which the general public is not familiar. Questions about life and death would need to be discussed in depth to come up with an appropriate answer.

3. Maximization of QALY as criteria of social wellbeing

We will now move on to the matter of the social value of QALYs, or in other words the problems we encounter when we must evaluate different ways of adding QALYs. Maximization of QALYs, which is what is behind the use of thresholds, generates some very important problems. In fact, the *Patient Protection and Affordable Care Act* (popularly known as Oba-

macare) explicitly prohibits the use of QALYs and thresholds[10,11] for an ethical reason, the double jeopardy argument, which considers that QALY maximization discriminates against patients with disabilities (Harris, 1987; Singer *et al.*, 1995; Harris, 1995). Proposals have been made to prevent this using QALYs (Nord *et al.*, 1999; Basu *et al.*, 2020), which although receiving a fair amount of academic attention (measured by the number of citations) have received no real practical attention. The second argument against QALY maximization is that of the initial severity. It is argued that the social value of a medical technology does not depend solely on the QALY gain but on the patients' starting point. Again, there have been various academic proposals (Cuadras-Morató *et al.*, 2001; Bleichrodt *et al.*, 2002; Stolk *et al.*, 2004) to incorporate this element into public decision making. We believe that unlike the double jeopardy problem severity can be effectively incorporated into the cost-utility model, among other reasons because it is easy for politicians to justify (Magnussen *et al.*, 2015; Svensson *et al.*, 2015). Furthermore, there are other reasons why QALY maximization may not maximize social welfare[12] that have been around for quite a long time. By way of example, Torrance (1973) asked, "Is it equally good, as the model assumes, to extend one life by a thousand days or a thousand lives by one day?". Such important questions as this are answered in the affirmative when we maximize QALYs, despite the lack of hard empirical evidence to this effect.

Last, as much as we can adjust the model for equity factors there is no way to avoid the controversy stemming from the implicitly utilitarist nature of QALYs, and especially the difference of opinion over the QALY model's adjustment for duration. By way of example, research focused on the criteria to treat terminally ill patients (Pinto-Prades *et al.*, 2014; McHugh *et al.*, 2019) suggests that a large part of the population considers that these patients' treatment should be evaluated using non-utilitarist criteria. In these cases, it is easier to simply accept the limitations of QALYs than to try to find alternatives based on complex weightings.

4. NICE *vs.* IPT revisited

Now let's go back to the beginning. Is it better to fund Eluxadoline given that its cost per QALY is so far below the 30,000-pound threshold, or is it better not to fund this medication as the IPT sustains? In real life, given the limitations of decision-making processes this question cannot be given a "scientific" answer. What we would need to know to be able to provide an adequate response is how each of the assumptions of the QALY model, both at the individual and at the social level, have influenced the final ratio. For example, is risk neutrality a good assumption in this case? Is independence of disjoint health states a good assumption in this case? Since it is not possible in real life to test the validity of those assumptions in every case, they are implicitly accepted as valid by default, which is why most decision makers would consider irrational not funding a medication with a cost per QALY that is 40% below the theoretical threshold. The only reason we might find ourselves questioning this decision is because there is a different jurisdiction that takes the opposite decision using different principles. How is it possible that the Spanish regulator rejects a technology which a cost

per QALY so low? It is this contradiction what migh question the model. It may be that this very low cost per QALY ratio does not reflect a high social value because the assumption that many small health improvements is the equivalent to fewer large improvements does not hold. In fact, the IPT suggests that this medication does little to improve health. If the social value is not the simple sum of QALYs, a cost per QALY of 12,000 pounds may not indicate a good allocation of resources. We are not suggesting that this is the reason why the NICE and IPT recommendations are opposing. What we are pointing out is that: a) there are multiple reasons why it may be perfectly rational not to fund a medication with a cost per QALY clearly below the threshold; and b) these reasons are not apparent if we only evaluate each medication in isolation in terms of the cost per QALY and no alternative viewpoints are examined.

We are not suggesting that QALY should not be used. What we do believe is that decisions as to whether or not to fund a medication or the price of health technologies should not be taken solely on the cost per QALY. Given the limitations of the QALY model, ways to make decisions that complement the QALY model must be sought. We can learn from other areas where researchers have encountered similar problems. Two examples are Sunstein *et al.* (2002) for the case of evaluating monetary of health injuries, and Kahneman's proposal (2009) to evaluate states of health using methods that take the imprecision in preferences into account.

Given that QALY maximization is no guarantee of having found the social "optimum", we currently have no choice but to rely on deliberative methods that compare the reasons for one decision or another in the simplest way. Economic evaluation cannot be sold as a foolproof instrument to come up with a single solution to a complex problem. Although it is not possible to produce a detailed explanation of those "deliberative methods" here, we do want to give an example of what we have in mind.

The Oregon Health Evidence Review Commission (HERC), after a failed attempt to prioritize health services using cost per QALY, has developed a "simple" method to prioritize health services. They apply the following algorithm

Health care category weight	×	+ Impact on Healthy Life + Impact on Suffering + Population Effects + Vulnerability of Population Affected + Tertiary Prevention	×	Effectiveness	×	Need for Service

The HERC explains that, *after deliberation*, some adjustments were made by hand for cases where the ranking given by this methodology did not reflect the importance of the service. Are we suggesting that this is a better method than the cost per QALY? No, that is not our main point. Our argument is that the "best" method to take decisions based on constructivist rationality does not exist. For this reason, it would be better to take decisions based on the comparison of different approaches that might produce different decisions. Those potential "contradictions" will very useful to compare different perspectives and values and

to question (if necessary) decisions based in the argument that "it is below the cost per QALY threshold".

The problem with the cost per QALY is that it is very difficult, if not impossible, to debate about the reasons behind the numbers. At the end of the day, the main argument is going to be that the cost per QALY falls within the range of "acceptable values" and, in practice, there is no way we can know the extent to which the result depends on the (sometimes arbitrary) assumptions of the model[13]. In contrast, simple models like the algorithm of the HERC given above have the advantage that the link between the assumptions of the algorithm and the decisions that it generates can be easily understood. Again, let us emphasize that our point is not that the HERC algorithm is better than the cost per QALY. Our main point is that we do not think that we can talk about the best or optimal method. For this reason, decisions should be based on the comparison of different approaches based on assumptions and values that can be explained and debated openly.

This does not exclude the possibility of improving methods to calculate QALYs but we doubt that research is going to produce the "gold standard" method. This is why it is unsurprising that different jurisdictions take different decisions based on the same evidence. To this effect, the fact that an IPT produces a recommendation that is different from the NICE report should be considered as an opportunity to re-think the logic of difficult decisions. Similarly, the fact that different regional governments disagree on the appropriacy of including certain services in their portfolio is to be expected because on many cases no single method will provide a clearly defined answer.

5. On the derivation of the threshold and the assumable QALY cost

The practical application of QALYs in prioritizing health today revolves largely around the cost-effectiveness thresholds of treatments and their transferal to policy adoption. As seen before and pointed out in other works (López-Casasnovas, 2019), the methodology used for the calculations is questionable, and so the results obtained must be taken cautiously and never linked with rules of thumb in matter as paramount as social cohesion. Furthermore, its indiscriminate application can constitute the basis of a bad management and a worse application of decision-making procedures in specific cases involving treatments, recipients, and guidelines.

The situation we have described has to do with: (i) the difficulty in effectively estimating a social evaluation beyond the perspective of the health, public and administrative budgetary systems. Greater ease of calculation cannot subrogate a social evaluation, especially in interstitial treatments with indirect social impacts and joint long-term implications; (ii) the fact that aspects of equity and the introduction of social values are shied away from regarding treatments in a debate nowadays limited to their identification (Cockson *et al.*, 2021, or the so-called CEA -cost effectiveness-distributive analysis). We therefore have cost-effectiveness situations that are harmful to equity and effective ones that improve it. Their outcomes must be subject not to an algorithm but to an explicit deliberation process; (iii) how to incorporate

non-health aspects of consumption, which in a public health system stem from a redistributive funding based on some given resources, into the evaluation, from the real and taxed income of those who elicit preferences. Opportunity cost considerations of each individual (and therefore distributive aspects) in non-health consumption are still lacking today; and (iv) anticipating a *threshold*, facilitating a basically concealed relationship between value and price close to it, which again allows the excess of the producer to be as high as possible below the maximum ceiling: In short, the opportunist behavior of the agents (pharmaceutical companies), meaning that given that they have prior knowledge of the threshold, they may propose (and assign) a price to keep the ratio just below the threshold or, in other words, the maximum price the decision-maker public is willing to pay.

In brief, one either believes in competitive markets (including here the bilateral monopoly) or in well thought out regulation and subjugating the markets to the common good. The two things at the same time have so far proved to be incompatible. Because one thing, for example, is to assess a new technology to be incorporated into the system subject to an economist cost criterion, and another is to use the economic cost analysis to restrict that technology to those who can generate more QALYs[14].

6. Calculation of the cost-effectiveness thresholds

Estimates of willingness to pay and the opportunity cost of healthcare resources are needed. They can assessed as 1) opportunity costs in terms of health foregone when costs fall on health care budgets -so called k threshold- and 2) opportunity costs in terms of foregone consumption (the "consumption value of health") when additional costs fall on consumption opportunities outside health care, or v threshold. The first is an issue of "fact," resulting from limits in the overall collective budget available for health care or constraints on the health system's abilities to increase expenditure. It reflects the health generated at present from the health care system (or that could be gained if expenditure were increased) and, therefore, reflects the "supply side" of the system. The second is an issue of "value" and depends on how individuals and society value health has compared with other forms of consumption or publicly funded non-health goods. This indicates what individuals and society want from the health care system, or the "demand side[15]."

Without wishing to get embroiled in the *welfarists-non welfarists* debate[16], what is certain is that the issue of whether budgetary spending is a faithful reflection of the social willingness to pay is a very controversial one. Budgets are the result of central public financial constraints plus regional variations in spending, and as such are very changeable depending on the period of the analysis. It is not, therefore, a commitment to pay for better health from which an implicit threshold for the inclusion or exclusion of a certain treatment from the portfolio of services can be derived. The elements of cost effectiveness underpinning the studies reviewed (Claxton *et al.*, 2015) identify cost in terms of expenditure, and effectiveness with life expectancy. Aggregation misleads the ratio and consequently the estimation cannot fully neutralize cross-cutting and temporary variations (Lakdawalla and Phelps, 2021).

The pretty much explicit objective of calculations of this type is to evaluate the marginal benefits of health care in terms of health, or in other words a kind of analysis of how the elasticities of health spending manifest in health outcomes. We note that with aggregated data this is the equivalent of relating increases in life expectancy (observed as a year and a half of life gained per decade) with the healthcare bill. If a stable ratio of heath care costs is maintained in terms of GDP, the threshold will necessarily vary in line with increases in income or the tax revenue collected per capita. To this effect, the threshold calculated is not going to be greatly different from the per capita income of the country.[17] The fine tuning of adjustments when estimating elasticities of this sort using other covariables and life expectancy adjusted by quality to approximate the cost per QALY hides the basic hypothesis implicit in the aggregated calculations.

The v-*thresholds* are thereby derived from budgetary data, associating cost with health outcomes. This is quite the opposite to what we health economists often say: it is not the quantity of the cost but its composition that matters and, in any case, 'more is not always better' and neither does "the best" depend solely on "more": better a bit of health in all the policies.

Another point to consider is that the regional budgets in Spain are not exogenous[18] and their variations (which are extremely rare beyond the incomes that enable state-funded healthcare) do not appear to contribute greatly to differences in health results (supposedly, replacing some treatments with other more cost-effective ones). Reverse causality, between costs and health, and health and costs, requires using instrumental variables which, as neat as they are, do not cover other deficits of the estimation (misspecification errors, heteroscedasticity, and others).

Last, in countries like ours where the data has serious limitations, going from resources to results (differences in life expectancy), to quality of life related to health outcomes in estimating QALYs through adjustments based on self-perceived health in surveys with wide age bands, generates risky leaps of faith in the conclusions. In principle, the theory of thresholds is based on identifying the displaced cost of treatments. This identification is always difficult. It affects certain treatments, and their recommended alternative uses. Resorting to cost increases, and not to replacing treatments equal in cost, does not serve the same methodological purpose (calculation of the opportunity costs)[19]. In both scenarios, it is very likely that changes will be restricted by the available human resources and some management limitations. The calculated figures cannot then be considered as socially acceptable opportunity costs, with concurrence with the CBA (cost-benefit analysis) theory.

7. Transfering the results of some studies to policymaking is not an easy task

If we ignore administrative red tape (for the *undo* option versus the do option), from the perspective of opportunity costs (the incremental health gained) full transferability of resources can be assumed, replacing less cost-effective services with others with a higher

ratio. It is postulated that the option *undo* without *do* is unthinkable. This logic is different from considering that if a treatment does not go beyond a threshold (say above 30,000 euros per QALY, or pounds in the UK) it is not worth funding. At best, this way of thinking represents the expected average marginal effect of incorporating new treatments, not the value of the existing ones in view of how the system has been funded. Neither does it consider the best possible alternative for combining resources for funded services in terms of treatments and beneficiary groups. Moreover, it ignores the expected entitlement of patients to access some services, including those on waiting lists for the conventional treatments included in the universal services catalogue. It is also insensitive to the fact that the QALY maximization target is not 'free' but is socially restricted by having to comply with certain rights or ethical principles such as not discriminating by gender, age, or origin. The threshold would be limited at the time of admitting new treatments without referencing a type of illness or patient. Furthermore, this way of reallocating resources may vary with changes in the budgetary constraint, which could be relaxed with an injection of revenues from specific taxes in favor of certain treatments and beneficiaries, and perhaps with co-payments according to income. Thresholds would then change too.

8. In any case, it appears that in practice there is no maximum QALY value

In effect, because of the severity of some health states, higher thresholds are sometimes implicitly accepted, possibly due to an increased sensitivity in terms of social disposition to fund serious health problems. This would determine higher thresholds for these situations, which would go against the intuition of continuous and decreasing marginal QALY utilities, thereby accepting treatment thresholds with higher ratios for worse initial states of health.

Furthermore, initial health outcomes can have different interpretations depending on the cause of the deterioration in health and distributive perceptions, for instance on whether the health problem randomly appeared or if it was the to be the expected outcome of a certain lifestyle or diet, or whether the individual may have had access to health services or not, all adding up the different elements of individual responsibility that intervene.

And last, committing to a threshold made known to new treatment providers may not appear to be the best possible way to manage pre-contractual opportunism in terms of incentive.

9. When the QALY is diluted by the relative effectiveness and the context-specific decision 'silos' are created

As Sunstein *et al.* (2001) points out in 'Predictability incoherent judgements', isolated rational decisions cease to be so when they are analyzed in specific contexts such as when the moral result invalidates the rational process. Treating those in a similar situation in the

same way leads to the possibility of considering silos in funding. This may be the case among all patients with certain pathologies, neoplasms, and rare diseases with a similar chance of survival. The collective funding effort cannot be effectively compared with situations that represent lower QALY costs for common treatments.

Comparing inequivalent situations not in terms of evaluating their impact but in the context of oncological, end of life, or "isolated" treatment processes can point to recommending that funding should be settled based on social disposition, paying for treatments that respond to specific situations and changing circumstances; and what is more ensuring that within that situation relative efficiency manages to prioritize those for whom the best results can be achieved. This concurs with the more general idea that public policy should not maximize QALY indiscriminately but with restrictions, be it the restriction of "a fair innings" for everyone under the veil of ignorance, or that of not all patients competing for the same resources. Clearly, the horizontal QALY cost cutting across the different care actions is minimally coherent, though informative, and points to the relative effort of inter-silos, as opposed to the relative proficiency that would prevail with intra-silo. It is a matter of "normalizing," or "relativizing" within a specific category. In the economy of damages and its punishment not all legal decisions are commensurate solely with the size of the damage. Other aspects to evaluate could be whether there is a replacement for the damaged entity (species in danger of extinction) or any negative externalities with respect to other situations. As Sunstein *et al.* (op. cit.) point out, coherence is important because it requires a minimum amount of rationality. However, this coherence must not be deceptive. 'Incoherence' can be better than perversely unjust coherence in the eyes of the community. Institutional solutions that could end up being seen as rational if viewed in isolation are morally unacceptable and must be prevented. We need then to seek decision-making "frameworks" or schemes that incorporate contexts that ensure a pre-established legal standard which is overall accepted by society.

10. Conclusions

It is a known fact that what is best is sometimes the enemy of what is good, and this is especially true if the good sought through a general application overlaps with the cost of adhering to the theoretical principle. Without a doubt, QALYs have been a huge advance in economic evaluation, their use imposing order of some kind on the prioritization of health services. Us health economists hold on so fast to results for economically evaluating cost effectiveness because we know the void that would be left in their absence. However, from a theoretical point of view, its limitations can neither be ignored nor its indiscriminate use sanctified. At the very least, without deviating from economic orthodoxy, economic evaluation should incorporate elements of equity such as initial levels of severity. Obtaining empirical evidence with this objective in mind is for us a priority, even though it is probably not enough to overcome the limitations of QALYs. Consequently, we suggest that economic evaluation must co-exist with methods and instruments that are maybe less sophisticated than QALYs but allow relative comparisons of health technologies to be made in a more simple and ho-

listic manner. There is no single way to take rational decisions about resource allocation and there is probably no single "optimum" allocation of health resources. One well-known way is to determine specific funds, or silos, based on social disposition to fund groups of treatments for similar pathologies and later use QALYs to internally order their relative efficiency and, for the purpose of social review, to derive the differences in their cost effectiveness thresholds.

The reflection made in this paper sets some limitations and throws up some uncertainties that must preach caution in terms of the blanket use of an instrument which, taken to extremes, can damage social cohesion and destroy the instrumental tradition of economic rationality in the discipline of the Economics of Health in this much valued area for the collective wellbeing.

Notes

1. Like in other areas, in the economic evaluation of health technologies, the Cost-Benefit Analysis provides reference values (e. g., the statistical value of life, the monetary value of time, the monetary value of reducing CO_2 emissions, etc.) that guide the public decisions which, in the area of health, are called "thresholds".
2. We are referring to the fact that funding and prioritising decisions in the health services have a multidisciplinary component that requires a more complex approach than the application of simple rules based on the maximization of a unique variable.
3. ICER = Incremental Cost Effectiveness Ratio.
4. Revealed preference methods refer to a range of valuation techniques which all make use of the fact that many non-market goods and services are implicitly traded in markets like in the case of environmental goods, risks, and time.
5. Stated preference methods are based on hypothetical scenarios where subjects are asked to value objects with several attributes (using methods like Contingent Valuation, Matching, Choice Experiments, etc.). The final objective of these studies is to estimate the Marginal Relation of Substitution between goods and/or attributes.
6. "When would the assumption of additivity not hold? It would not hold if the law of diminishing returns applied-for example, if being in state S_D for 10 days were not 10 times as bad as being in state S_D for 1 day, or if a transition from state S_F to S_C for one person has less value than a transition from state S_F to S_E for one person, plus a transition from S_E to S_D for another, plus a transition from S_D to S_C for a third." 1041.
7. MUI: the utility of a health state A is always the same fraction of the utility of full health for all time horizons.
8. CPT: the proportion of remaining life that one would be willing to trade-off for a specified quality improvement is independent of the amount of remaining life.
9. Risk Aversion parameter is 1.
10. "The Secretary shall not use evidence or findings from comparative clinical effectiveness research conducted under section 1181 in determining coverage, reimbursement, or incentive programs under title XVIII in a manner that treats extending the life of an elderly, disabled, or terminally ill individual as of lower value than extending the life of an individual who is younger, nondisabled, or not terminally ill." (Section 1320e-1(c)).
11. "The Patient-Centered Outcomes Research Institute established under section 1181(b)(1) shall not develop or employ a dollars- per-quality adjusted life year (or similar measure that discounts the value of a life because of an individual's disability) as a threshold to establish what type of health care is cost effective or recommended". (Section 1320e–1(e)).

12. Oregon Health Authority (2021) Prioritization of Health Services A Report to the Governor and the 81st Oregon Legislature. Available at https:\www.oregon.gov\oha\HPA\DSI-HERC\Documents\2021-Biennial-Report-to-Governor-and-Legislature.pdf.

13. Would we have obtained different results if we had used the Health Utility Index instead of the Euroqol, the Standard Gamble instead of the Time Trad-Off, a 20-year duration instead of 10 years in the Time Trade-Off, Choice Experiments instead of Time Trade-Off, a different assumption about the utility function for life years? and many others.

14. Lakdawalla and Phelps developed this theory for the 'Health Technology assessment with diminishing returns to health' through the instrument named GRACE (Generalized Risk Adjusted Cost Effectiveness, en Value in Health 24 (2).

15. For economic evaluation it is important to consider what type of opportunity costs would result from investment in new activities. If opportunity costs result in the form of health forgone (e. g., through displacement of other health-generating interventions), then the cost effectiveness threshold should reflect this (the "k," the amount of money that would displace one QALY's worth of health care investment). If opportunity costs are in terms of other forms of consumption, then the cost effectiveness threshold should reflect the consumption value of health (denoted as "v"). See Woods, B. *et al*., 2016.

16. See Sakowsky, R. A., 2021.

17. With another words, holding public health care expenditure constant (say at 7%), a 10% increase of GDP over a decade would imply a 10% increase of PHCE. At present, this increase goes together with an increase of, say,1 year of life expectancy for the decade. In real terms, GDP per capita will give the value of this extended year per capita. Simplifying the calculation, if each year we assign 7% of the per capita income to health, in 10 years' time we will have spent €21 thousands. This would go with an increase of life expectancy of one extra year of life. It would be apparent that willingness to pay for an extra year is 21 thousand euros.

18. In the case of Spain, on using the data territorial spending is ignoring that fact that from the start funding has depended on cost effectiveness, which does not allow its exogeneity to be assumed, and neither does it depend on situational budget cuts. This is the same as recognising that spending does not follow a health care pattern since funding is allocated based on spending needs according to the population, ageing, weight of equivalent beneficiaries, etc.

19. Recent works looking for k-thresholds by researchers at Erasmus University Rotterdam, and the more global research of Cutler *et al*. (2020) for health outcomes from health care innovations for the USA, which have focused on certain treatments for which the differences in results are identifiable and have less associated confounding elements etc., offer in our view a more solid approach in terms of methodological robustness.

References

Abellán, J. M., Herrero, C. and Pinto-Prades, J. L. (2016), "QALY-Based Cost Effectiveness Analysis", *The Oxford Handbook of Well-Being and Public Policy*, chapter 6: 160-192, Oxford UP.

Agencia Española de Medicamentos y Productos Sanitarios (2019), "Informe de Posicionamiento Terapéutico e Eluxadolina (Truberzi ®) en el tratamiento del Síndrome del Intestino irritable con diarrea", IPT, (61/2019). V1. Availabe at: https://www.aemps.gob.es/medicamentosUsoHumano/informesPublicos/docs/IPT- eluxadolina-Truberzi.pdf?x74012.

Augestad, L. A., Stavem, K., Kristiansen, I. S., Samuelsen, C. H. and Rand-Hendriksen, K. (2016), "Influenced from the start: anchoring bias in time trade-off valuations", *Quality of Life Research*, 25(9); 2179-2191.

Basu, A., Carlson, J. and Veenstra, D. (2020), "Health Years in Total: A New Health Objective Function for Cost-Effectiveness Analysis", *Value in Health*, 23(1): 96-103.

Bleichrodt, H., Herrero, C. and Pinto, J. L. (2002), "A proposal to solve the comparability problem in cost-utility analysis", *Journal of Health Economics*, 21(3): 397-403.

Bleichrodt, H. and Pinto Prades, J. L. (2009), "New evidence of preference reversals in health utility measurement", *Health Economics*, 18(6): 713-726.

Bleichrodt, H., Pinto-Prades, J. L. and Abellán-Perpiñán, J. M. (2003), "A consistency test of the time trade-off", *Journal of Health Economics*, 22(6): 1037-1052.

Bush, J. W., Fanshel, S. and Chen, M. M. (1972), "Analysis of a tuberculin testing program using a health status index", *Socio-economic Planning Sciences*, 6(1): 49-68.

Campillo-Artero, C. (2013), "La regulación de la prestación farmacéutica hoy en España: la opinión de GENESIS contrapunteada por tres voces independientes", *Economía y salud: boletín informativo*, (76): 23-30.

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S. and Sculpher, M. (2015), "Methods for the estimation of the NICE cost effectiveness threshold", *Health Technology Assessment*, 19(14): 1-504.

Cockson, R., Griffin, S., Norheim, O. F., Culyer, A. J. and Chalkidou, K. (2021), "Distributional Cost-Effectiveness Analysis Comes of Age", *Value in Health*, 24: 118-120.

Culter, D. M., Ghosh, K., Bondarenko, I., Messer, K. L., Stewart, S. T., Raghunathan, T. and Rosen, A. B. (2020), "Attributing medical spending to conditions: A comparison of methods", *Plos one*, 15(8): e0237082.

Cuadras-Morató, X., Pinto-Prades, J. L. and Abellán-Perpiñán, J. M. (2001), "Equity considerations in health care: the relevance of claims", *Health Econ*, 10(3): 187-205.

Devlin, N., Shah, K., Feng, Y. *et al*. (2018), "Valuing health-related quality of life: an EQ 5D 5L value set for England", *Health Economics*, 27(1): 7-22.

Fanshel, S. (1972), "A meaningful measure of health for epidemiology", *International Journal of Epidemiology*, 1(4): 319-337.

Fanshel, S. and Bush, J. W. (1970), "A health-status index and its application to health-services outcomes", *Operations research*, 18(6): 1021-1066.

Fanshel and Busch (1970), "When would the assumption of additivity not hold? It would not hold if the law of diminishing returns applied-for example, if being in state S_D for 10 days were not 10 times as bad as being in state S_D for 1 day, or if a transition from state S_F to S_C for one person has less value than a transition from state S_F to S_E for one person, plus a transition from S_E to S_D for another, plus a transition from S_D to S_C for a third." 1041.

Hanmer, J., Cherepanov, D., Palta, M., Kaplan, R. M., Feeny, D. and Fryback, D. G. (2016), "Health condition impacts in a nationally representative cross-sectional survey vary substantially by preference-based health index", *Medical Decision Making*, 36(2): 264-274.

Harris, J. (1995), "Double jeopardy and the veil of ignorance-a reply", *Journal of Medical Ethics*, 21(3): 151-157.

Harris, J. (1987), "QALYfying the value of life", *Journal of Medical Ethics*, 13(3): 117-123.

Hernández Álava, M., Wailoo, A. and Pudney, S. (2018), "Quality review of a proposed EQ-5D-5L value set for England", *EEPRU report*.

Kahneman, D. (2009), "A different approach to health state valuation", *Value in Health*, 12: S16-S17.

Lakdawalla, D. N. and Phelps, C. E. (2021), "La *Health Technology assessment with diminishing returns to health* a través del instrumento denominado GRACE (generalized Risk Adjusted Cost Effectiveness)", *Value in Health*, 24(2).

López-Casasnovas, G. (2019), "A vueltas con los umbrales coste efectividad", *CRES Health Policy Paper* (2019/1). Available at https://repositori.upf.edu/bitstream/handle/10230/36690/Lopez-Casanovas_HealthPolicyPaper-1.pdf?sequence=1&isAllowed=y.

McHugh, N., Pinto-Prades, J. L., Baker, R., Mason, H. and Donaldson, C. (2019), "Exploring the relative value of end-of-life QALYs: Are the comparators important", *Soc Sci Med*, 245: 112660.

Magnussen, J., Aaserud, M., Granaas, T., Magelssen, M., Syse, A., Celius, E. G. *et al*. (2015), På ramme alvor-Alvorlighet og prioritering. Department of Health (p. 71). English summary available from: https://www.regjeringen.no/contentassets/d5da48ca5d1a4b128c72fc5daa3b4fd8/summary_the_magnussen_report_on_severity.pdf.

National Institute for Health and Care Excellence (2017), Eluxadoline for treating irritable bowel syndrome with diarrhoea (NICE Guideline). Available at www.nice.org.uk/guidance/ta471.

Nord, E., Pinto, J. L., Richardson, J., Menzel, P. and Ubel, P. (1999), "Incorporating societal concerns for fairness in numerical valuations of health programmes", *Health Economics*, 8(1): 25-39.

Pinto-Prades, J. L., Attema, A. and Sánchez-Martínez, F. I. (2019), "Measuring health utility in economics", in *Oxford Research Encyclopedia of Economics and Finance*.

Pinto-Prades, J. L., McHugh, N., Donaldson, C. AND Manoukian, S. (2019), "Sequence effects in time trade-off valuation of hypothetical health states", *Health Economics*, 28(11): 1308-1319.

Pinto-Prades, J. L., Sánchez-Martínez, F. I., Abellán-Perpiñán, J. M. and Martínez-Pérez, J. E. (2018), "Reducing preference reversals: The role of preference imprecision and nontransparent methods", *Health Economics*, 27(8): 1230-1246.

Pinto-Prades, J. L., Sánchez-Martínez, F. I., Corbacho, B. and Baker, R. (2014), "Valuing QALYs at the end of life", *Social Science & Medicine*, 113: 5-14.

Pliskin, J. S., Shepard, D. S. and Weinstein, M. C. (1980), "Utility functions for life years and health status", *Operations research*, 28(1): 206-224.

Sakowsky, R. A. (2021), "Disentangling the welfarism/extra-welfarism distinction: Towards a more fine-grained categorization", *Health Economics*, 2021(30): 2307-2311.

Singer, P., McKie, J., Kuhse, H. and Richardson, J. (1995), "Double Jeopardy and the Use of QALYs in Health Care Allocation", *Journal of Medical Ethics*, 21(3): 144-150.

Sugden, R. (2021), "How Hayekian is Sunstein's behavioral economics?", *Behavioural Public Policy*, 1-10.

Sunstein, C. R., Kahneman, D., Schkade, D. and Ritov, I. (2002), "Predictably Incoherent Judgments", *Stanford Law Review*, 54(6),: 1153-1215.

Stolk, E. A., van Donselaar, G., Brouwer, W. B. F. and Busschbach, J. J. V. (2004), "Reconciliation of economic concerns and health policy: illustration of an equity adjustment procedure using proportional shortfall", *Pharmacoeconomics*, 22(17): 1097-1107.

Svensson, M., Nilsson, F. O. and Arnberg, K. (2015), "Reimbursement Decisions for Pharmaceuticals in Sweden: The Impact of Disease Severity and Cost Effectiveness", *Pharmacoeconomics*, 33(11): 1229-1236.

Taylor, M., Chilton, S., Ronaldson, S., Metcalf, H. and Nielsen, J. S. (2017), "Comparing increments in utility of health: an individual-based approach", *Value in Health*, 20(2): 224-229.

Torrance, G. W. (1973), "Health index and utility models: some thorny issues", *Health services research*, 8(1): 12.

Torrance, G. W., Thomas, W. H. and Sackett, D. L. (1972), "A utility maximization model for evaluation of health care programs", *Health services research*, 7(2): 118-133.

Van Baal, P., Perry-Duxbury, M., Bakx, P., Versteegh, M., van Doorslaer, E. and Brouwer, W. (2018), "A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending", *Health Economics*. 1-14.

Wakker, P. P. (2008), "Lessons learned by (from?) an economist working in medical decision making", *Medical Decision Making*, 28(5): 690-698.

Woods, B., Revill, P., Sculpher, M. and Claxton, K. (2016), "Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research", *Value in Health*, 2016 Dec, 19(8): 929-935.

Zhou, T., Chen, Z., Li, H. and Xie, F. (2021), "Using Published Health Utilities in Cost-Utility Analyses: Discrepancies and Issues in Cardiovascular Disease", *Medical Decision Making*, 0272989X211004532.

Resumen

Los Años de Vida Ajustados por Calidad (AVAC) se han consolidado como medida paradigmática de los resultados en salud, siendo amplia su utilización en la evaluación económica, y con pretensiones de emulación en otros ámbitos de la política pública. Este texto tiene como objetivo ordenar los argumentos que subyacen en la utilización de los AVAC en los estudios de coste efectividad y que condicionan así su aplicabilidad. Se reacciona de este modo a un uso a menudo poco razonado en lo que atañe a sus limitaciones a la hora de priorizar las prestaciones sanitarias. En otras palabras, los autores se muestran con matizaciones a favor del instrumento, pero en contra en todo caso de sus entusiastas.

Palabras clave: AVAC, coste efectividad, umbrales, evaluación económica, farmaeconomía.

Clasificación JEL: H51, I31, I18, K23, I14.