

A Criticism of Healthy-years Equivalents

PETER WAKKER, PhD

The following questions describe the scope of this paper. When decision trees are used to analyze optimal decisions, should end nodes be evaluated on the basis of QALYs or on the basis of healthy-years equivalents? Which measures should be used in communications with others, e.g., patients? Which of these measures incorporate risk attitudes, and which do not? It is demonstrated that the healthy-years equivalent measure does not stand scrutiny. *Key words:* utility; QALY; quality of life; healthy-years equivalent; standard gamble. (Med Decis Making 1996;16:207-214)

In recent articles on medical decision making, a lively debate has arisen around the healthy-years equivalents (HYES) advanced in 1989 by Mehez and Gafni.¹⁻⁹ A conclusion of this paper is that the claims of these authors regarding HYES do not stand scrutiny.

All the discussions of HYES in the literature have assumed the von Neumann-Morgenstern theory of expected utility, so this paper also does so in the main body of the text.¹⁰ Therefore, it first explains the general expected-utility theory. Quality-adjusted life years (QALYs) are described as a specified version of expected utility, and general principles for the choice between general and specified models are explained. "Counterexamples" and "preference reversals," advanced by Gafni et al. in defense of HYES, are based on no other argument than the simple fact that HYES are more general than QALYs. Then we turn to the comparison of HYES and time tradeoffs (TTOs). Several authors have explained that HYES yield the same results as TTO questions under expected utility theory. It is explained that Gafni et al. thought otherwise because of a misunderstanding of the riskless-value/risky-utility difference and of the "degeneracy principle" of expected utility. Ambiguity in Gafni and colleagues' writings about the risk-neutrality assumption for healthy life years has led to misunderstandings in criticisms of their work. Depending on the ambiguity, either HYES violate the empirical finding of risk aversion, or they are simply a restatement of general expected-utility theory where no information about risk attitudes is recorded in the HYES. Under violations of expected utility, HYES do not lead to alternative rationality theories because they apparently invoke violations of transitivity and continuity. Empirically, the presently

existing evidence suggests that HYE elicitation incorporate larger -measurement errors than do TTO elicitation.

General Expected-utility Theory

Outcomes are assumed to be health profiles in this paper. A health profile describes a time period, designating the life duration of the patient, and specifies for each time point the health state of the patient. An example is the health profile denoted by (hoarse, 5 yr + mute, 5 yr), where the patient lives for ten years, and during the first five years her voice is hoarse; during the last five years the patient is mute. A chronic health profile is one where the health state of the patient is the same until death. (Q, L) designates a chronic health state of L years of life in health state Q. In this paper we restrict attention to positive health states, i.e., for each health state we assume that longer life durations are always valued higher. This is not essential to the conclusions of the paper but simplifies some formulations.

Next we incorporate risk into the model. A typical notation is ((Q₁, 71, 0.4; Q₂, 41, 0.6) for a medical treatment that with probability 0.4 results in seven years of life in health state Q₁, and with probability 0.6 in four years of life in health state Q₂. The term *gamble* designates such probability distributions over outcomes, i.e., over health profiles. The notation for a general gamble is

$$(\mathbf{x}_1, \mathbf{p}_1; \dots; \mathbf{x}_n, \mathbf{p}_n)$$

where $\mathbf{p}_1, \dots, \mathbf{p}_n$ are the probabilities and the outcomes $\mathbf{x}_1, \dots, \mathbf{x}_n$ are health profiles.

"Degenerate gambles" are gambles that do not entail risk, but with certainty result in an outcome. For instance, the "gamble" ((mute, L), 1) with certainty results in a life duration of L years without speech. Such gambles are included in the domain of the theory (Keeney and Raiffa,¹¹ Formula 5.21. Because this is a central point in the discussion below, it is emphasized: **degeneracy principle of ex-**

Received March 11, 1994, from the Medical Decision Making Unit, University of Leiden, Leiden, The Netherlands. Revision accepted for publication November 14, 1995.

Address correspondence and reprint requests to Prof. Wakker: Medical Decision Making Unit, University of Leiden (AZL), P.O. Box 9600, Leiden, 2300 RC, The Netherlands.

pected utility. The von Neumann-Morgenstern expected-utility theory also governs choices between degenerate (“riskless”) gambles.

Under von Neumann-Morgenstern expected-utility theory, gambles are preferred according to their expected utilities:

$$p_1U(x_1) + \dots + p_nU(x_n)$$

Here, U denotes the utility function that values outcomes. Expected utility immediately implies that preferences are transitive, i.e., if a first gamble is preferred to a second, and the second to a third, then the first gamble is preferred to the third.

For most applications, the expected-utility model is considered too general, and specific assumptions are made about utility. For instance, utility is assumed to be a member of a parametric family so that only the parameters need to be assessed, or utility is assumed to be decomposable in some tractable manner. The next section discusses some specific assumptions concerning utility.

QALYs, and General versus Specified Models

This section shows how the QALY model can be obtained as a specification of expected utility. For the general QALY approach, two specifying assumptions are made for the utility U . These are:

1. Additive decomposability over disjoint time periods

That is to say, the utility $U(\text{hoarse, 5 yr} + \text{mute, 5 yr})$ can be written as a sum $U(\text{hoarse, 5 yr}) + U(\text{mute in years 6, . . . , 10})$, where the latter term describes the utility contribution of the years 6 to 10 in the health state mute.

2. Utility independence of life duration and health state

This condition refers only to utilities for chronic health profiles, and implies that utility can be written as a multiplicative form, i.e.,

$$U(Q, L) = W(Q) \cdot V(L)$$

where $V(L)$ depends only on the life duration L irrespective of the corresponding health state Q , and $W(Q)$ depends only on the health state Q irrespective

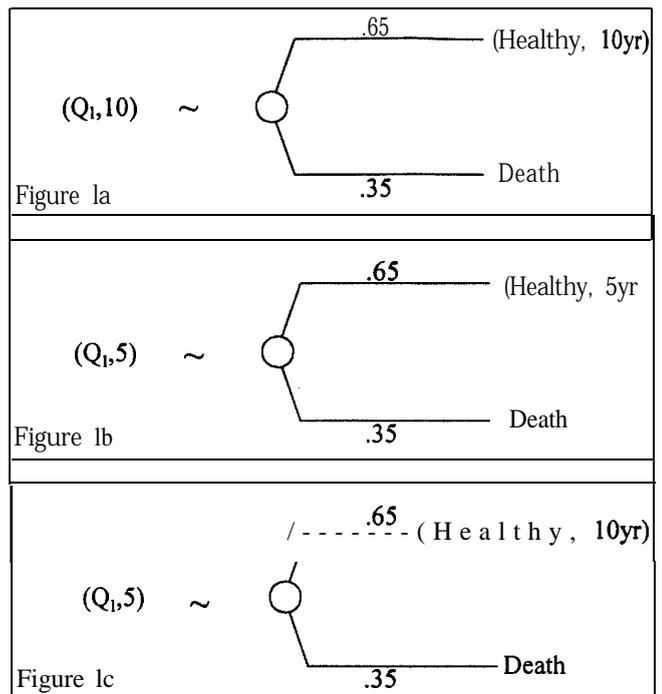


FIGURE 1. Utility independence. By utility independence, the indifference in figure 1b is implied by the indifference in figure 1a. The preference in figure 1c is discussed by Mehrez and Gafni.

of the life duration that is combined with.* By normalization, $V(M) = 1$ for the maximal life duration M , and $W(Q) = 1$ for the optimal health state $Q = \text{healthy}$.

An implication of utility independence for standard gambles is depicted in figures 1a and 1b: As maximum life duration, we take ten years; hence we set $W(10 \text{ yr}) = 1$. If in figure 1a indifference (denoted by \sim) results for the probability 0.65, suggesting that $W(Q_1) = 0.65$ for ‘state of health Q_1 ’, then in figure 1b indifference should also result for the probability 0.65. In other words, it should not matter for the elicitation of $W(Q_1)$ whether the health states are combined with a life duration of ten years or with a life duration of five years. (Figure 1c is used later, to discuss an example of Mehrez and Gafni.) The preference conditions that characterize the above specifications are discussed in further detail by Bleichrodt.¹²

The above two assumptions define the general QALY model as a specification of expected utility. In QALY calculations, a third assumption is usually added, i.e., it is usually assumed that all life years

*Formally, additive models $W(Q) + V(L)$ can also occur under utility independence. They, however, can not accommodate the special role of 0 years of life, where the health state should not matter, and therefore are not mentioned in the text.

have the same value, so that $V(\text{first year}) = \dots = V(\text{tenth year})$. This can be characterized by risk neutrality with respect to gambles for life duration. Then $V(L) = L$ can be taken for all life durations L (it is now more convenient to drop the normalization), and $U(\text{hoarse, 5 yr} + \text{mute, 5 yr})$ is easily calculated as

$$W(\text{hoarse}) \cdot 5 + W(\text{mute}) \cdot 5 \quad (1)$$

It is quite obvious that the additivity assumption over disjoint periods and the utility-independence assumption, as well as the risk-neutrality assumption, cannot be perfect empirical descriptions. Hence, even if a person perfectly well satisfies expected utility, then still the QALY calculations sometimes prescribe preferences contrary to the actual preferences. The QALY assumptions are at best satisfied approximately. The advantage of the assumptions is, however, that they greatly simplify the measurement requirements, because fewer measurements are needed to assess the utilities of general health profiles. For example, in the QALY model, utilities need be assessed for chronic health states only. They uniquely determine the utilities for all health profiles. Also, separability of utility simplifies the integration into Markov models. Whether the greater tractability of analysis outweighs the loss of empirical realism is a question that cannot be answered in a universal manner; the answer depends on context and application.

It should be understood that the trading off of greater tractability versus loss of realism is an issue in all empirical fields, and always occurs when a choice is to be made between more and less general models, or more and fewer free parameters.⁴ This issue is central in multiattribute utility, where special forms of functions are conjectured primarily for being tractable," or for the "nonexpected-utility" models in risk theory, where an increase of empirical realism is obtained as compared with expected utility, but in return many convenient methods of expected-utility analysis then must be abandoned. Hence, an argument for a more general model that emphasizes the gain in empirical realism only is not very interesting (compare Gafni and Zylak,¹³ second paragraph), as is a criticism that does nothing but discuss the loss in tractability. Similarly, hypothetical examples to demonstrate that preferences under the restrictive model can deviate from true preferences if the latter can be anything under the more general model are tautological and do not provide new evidence. Only when the pros and cons are considered jointly, in relation to the domain of application, can sensible choices be made between more and less general models.

Healthy-years Equivalents Defined and Discussed

This section presents the healthy-years equivalent (HYE) idea of Mehrez and Gafni,¹ and discusses it.

HYE, **the definition.** Mehrez and Gafni follow the approach of the main body of this paper by basing their analysis on expected utility. We continue denoting the utility function by U . The HYE for an outcome, e.g., (hoarse, 5 yr + mute, 5 yr), is defined as the number of years in perfect health that has the same U value as the outcome. For example, if

$$U(\text{hoarse, 5 yr} + \text{mute, 5 yr}) = U(\text{healthy, 9.5 yr}) \quad (2)$$

then

$$\text{HYE}(\text{hoarse, 5 yr} + \text{mute, 5 yr}) = 9.5 \text{ yr}$$

For measuring HYE, Mehrez and Gafni propose a two-stage scheme. Thus, from the two indifferences

$$\begin{aligned} &(\text{hoarse, 5 yr} + \text{mute, 5 yr}) \\ &\sim (\text{healthy, 10 yr}), 0.97; (\text{death}), 0.03 \quad (3) \end{aligned}$$

and

$$\begin{aligned} &(\text{healthy, 10 yr}), 0.97; (\text{death}), 0.03 \\ &\sim (\text{healthy, 9.5 yr}) \quad (4) \end{aligned}$$

they infer that $\text{HYE}(\text{hoarse, 5 yr} + \text{mute, 5 yr}) = 9.5$ yr. Indeed, as the right side of equation 3 is the same as the left side of equation 4, substitution of expected utility in the described indifferences implies that equation 2 holds. In general, for any health profile, Mehrez and Gafni in a first stage use a standard gamble to find a lottery over the best and the worst available outcomes that is indifferent to the health profile. Then, in a second stage, they elicit the number of years in full health (the HYE) indifferent to the lottery found in the first stage. This procedure was first described for chronic health profiles only,¹ then applied to general health profiles.¹⁴

Mehrez and Gafni advance as the main advantage of HYE that they do not depend on the restrictive assumptions from the QALY model. Measurement of HYE does not require an additive evaluation of disjoint time periods or utility independence, but invokes only the principles of expected utility. Mehrez and Gafni argue that HYE:

- are based on expected utility;
- incorporate risk attitudes;
- are fairly simple to measure; and
- have a meaning that is intuitively appealing.

We next turn to a first criticism of the HYE idea.

HYE questions are formally equivalent to TTO questions. A first criticism, expressed by several discussants of HYE, is that the measurement of HYE can be simplified.^{2,5,7,8,15-17} If one considers the indifference in equations 3 and 4, then by transitivity the indifference between (hoarse, 5 yr + mute, 5 yr) and (healthy, 9.5 yr) follows. It is easier to immediately ask for the number of years in perfect health that provide indifference to the health profile under consideration. Thus one immediately elicits the indifference

$$(hoarse, 5 yr + mute, 5 yr) \sim (health, 9.5 yr)$$

Such direct elicitation are standardly used in the TTO method. Thus, the criticism proceeds, the HYE measurement procedure that Mehrez and Gafni proposed is nothing but a roundabout manner for asking TTO questions.

In response to this criticism, Mehrez and Gafni⁶ present the following example. (The same example and argument were also published elsewhere.⁷ Assume the following three indifferences.

$$(dial.un., 15 yr) \sim (healthy, 15 yr), 0.7; (death), 0.3 \tag{5}$$

$$(healthy, 15 yr), 0.7; (death), 0.3 \sim (healthy, 8 yr) \tag{6}$$

$$(dial.un., 15 yr) \sim (healthy, 11 yr) \tag{7}$$

The first two indifferences reveal that

$$HYE(dial.un., 15 yr) = 8 yr$$

The third indifference reveals that

$$TTO(dial.un., 15 yr) = 11 yr$$

The theoretical analysis that Mehrez and Gafni propose for these choices is as follows. In choice situations of risk, the person conforms to expected utility, and the identity

$$U(dial.un., 15 yr) = U(healthy, 8 yr) \tag{8}$$

is revealed by the first two indifferences. In riskless situations, however, the person maximizes a “value” function V that may very well be different from the risk function U. So the third indifference reveals

the identity

$$V(dial.un., 15 yr) = V(healthy, 11 yr) \tag{9}$$

Mehrez and Gafni emphasize that their example is meant to fully agree with expected-utility theory. However, it does not. By the degeneracy principle, expected utility also governs choices between riskless options. If $U(dial.un., 15 yr) = U(healthy, 8 yr)$, then in the TTO question the person must express indifference between (dial.un., 15 yr) and (healthy, 8 yr) and the third indifference is excluded.

The error underlying Mehrez and Gafni’s ideas.

The conclusion of the following discussion was also obtained by Loomes.¹⁵ To justify the indifferences in equations 5, 6, and 7, leading to equations 8 and 9, Mehrez and Gafni invoke the “value-utility” theory that assumes a difference between risky utility functions and riskless value functions. The best-known papers about this theory may be those of Dyer and Sarin^{18,19}; indeed, Gafni et al.^{9,20} have invoked these references. Let me emphasize that the theory is entirely within the realm of expected utility.

As an example to illustrate the theory, assume that, for a maximum life duration of 25 years, the value function for healthy life years is $V(x yr) = x/25$; hence it is linear. The utility function for healthy life years is the square root of $V(x yr)$, i.e., $U(x yr) = \sqrt{(x/25)}$. Both functions assign value 0 to 0 years and value 1 to 25 years, and are monotonically related, i.e., they order the certain outcomes in the same way (longer life durations are preferred). The difference between the two functions is in the ordering of differences. In the riskless context, the V-difference between 25 years and 16 years is $V(25 yr) - V(16 yr) = 0.36$, which is the same as the V-difference between 9 years and 0 years. In the risky context, however, the U-difference between 25 years and 16 years is $U(25 yr) - U(16 yr) = 0.2$, which is smaller than 0.6, the U-difference between 9 years and 0 years.

The value-utility theory has been developed to distinguish, within the framework of expected utility, risk attitude from the riskless valuation of preference intensities. For instance, in the above example it can be claimed that the (riskless) valuation of preference intensities for life duration is linear so that there is constant marginal value, but still in risky decisions the utility function is not linear (linear = risk neutral) but concavely curved, in agreement with risk aversion. The value-utility theory does fully agree with expected utility, and does not permit violations of expected utility, as in Mehrez and Gafni’s example. A phenomenon such as that described in equations 8 and 9 is not permitted in this theory. U and V must order the riskless outcomes in the same way, and U is necessarily a monotonic transform of V; U and V are “ordinally identical.”^{11,18,19}

Gafni et al.^{6,9} used the terms “isovalue curve” and “isoutility curve” to designate sets where the value function, and the utility function, respectively, are constant. As we have just seen, isovalue and isoutility curves *must* be *identical*. The most-studied case of the value-utility theory is where U is a power of V .²¹⁻²⁵ The possibility that two functions can order riskless outcomes in the same manner and still be different has not always been recognized by Gafni et al.†

In summary, value and utility are distinct because they order *differences of* riskless outcomes differently, rather than riskless outcomes themselves.

Other problems for Mehrez and Gafni's distinction between risky and riskless choices. It has been demonstrated above that the theory of Mehrez and Gafni, distinguishing between risky utility and riskless value in an ordinal manner, deviates from expected-utility theory. That by itself does not necessarily mean that their theory is without interest. However, on closer study, their theory violates basic principles of rationality. As pointed out by many, choices such as those in equations 5, 6, and 7 constitute a violation of transitivity. A second problem is that the theory leads to violations of continuity, i.e., if one gradually changes a risky choice situation into a riskless one, then in the limit as the amount of risk approaches 0, counterintuitive preference reversals can occur. For brevity I do not elaborate on this claim.

HYEs do not incorporate risk attitudes. It will now be demonstrated that the risk attitude that is indeed inherent in the first stage of the HYE-measurement procedure exactly cancels out in the second stage.^{2,5} Gafni et al. contest that claim.^{3,6,9} Assume, to that effect, that two decision makers order riskless outcomes in the same manner, but have different risk attitudes. Because TTO questions involve only riskless outcomes, the two decision makers then provide the same replies to the TTO questions. We saw above that the replies to TTO questions coincide with HYE; therefore, their HYE values are also identical. Thus, HYE cannot, for a given ordering of riskless outcomes, distinguish between different risk attitudes, and any effect of risk attitude in the first stage of HYE measurement must be neutralized in the second stage. Let us, to avoid any ambiguity, consider some mathematical conclusions that can be drawn at this stage.

THEOREM 1

Under expected utility, HYE values are identical to TTO replies. Consequently, if two expected-utility maximizers have the same riskless preferences but have different risk attitudes (i.e., their preferences in gambles differ), then they have the same HYE values. Hence, HYE values do not incorporate risk attitudes.

The above theorem shows that TTOs and HYE values do not incorporate risk attitudes, and that risk attitudes cannot be derived from them without additional information. This is contrary to the claims of Gafni et al.' The main assumption of expected utility underlying the above reasoning (besides the domain assumptions entailing the degeneracy principle) was transitivity of preference, and the specific integral form of expected utility was not used. Hence the TTO and HYE values are not only independent of the risk attitudes of persons under expected utility, but, even stronger, they do not depend on expected utility, and also hold for transitive “non-expected-utility” generalizations of expected utility. A similar observation was made by Loomes.¹⁵ This observation provides negative answers to the research questions proposed at the ends of some papers.^{3,6,9}

HYEs cannot be used in decision-tree calculations. Another issue concerns the question whether HYE values are meant to be used in decision analyses. That is, can the expected number of HYE values be an index for evaluating medical treatments, as the expected number of QALYs evaluates medical treatments in the QALY model? This implies risk neutrality for healthy life years, as the HYE of any health profile (L y-r, healthy) is L . Gafni et al. write in many places that HYE values fully represent individuals' preferences, and that HYE values should be used instead of QALYs or utilities. Studies in diagnostic radiology have used HYE values in this way to analyze the use of ionic versus nonionic contrast media.” In several of their publications, Gafni et al. have suggested that they indeed intend HYE values to be used in this manner (“equate the HYE with utility”).^{1,14}

Based on the mentioned works of Gafni et al., discussants have criticized HYE values for the assumed risk neutrality for healthy years.^{5,15,16,17} However, Mehrez and Gafni contradict the claim⁵ that the HYE approach assumes risk neutrality for healthy life years. Mehrez and Gafni⁶ (equation 1) state that HYE values as such should not be used in an expected-utility analysis, but should be transformed into utilities and only then used to calculate expectation. Gafni²⁷ criticizes QALYs with risk neutrality for violating the commonly-found risk aversion, although this criticism seems to be directed to risk neutrality for nonperfect-health years only. Confusion resulting

†This occurs most clearly in Mehrez and Gafni⁶ (page 288, next to last paragraph). See also Mehrez and Gafni' (equation 4; the relevant point is properly mentioned in Footnote ‡ there) and Gafni et al.⁹ (page 331, lines 13–14). Discussions in Mehrez and Gafni' (page 147, second paragraph) and Mehrez and Gafni²⁰ (page 1282) do point out that cardinal (rather than ordinal) measures of outcome are relevant, but use this to criticize QALYs and TTOs, respectively, and not HYE values.

from the two interpretations of HYE (with and without risk neutrality for healthy years) underlies other criticisms.^{7,8,15 ‡}

The version of HYE theory that transforms HYE into utility values simply reduces (after deletion of its erroneous parts) to the general expected utility as described above, where no restrictive QALY-like assumption has been added. In other words, this version of the HYE theory has nothing new to offer and should not be given a new name. The version of HYE theory that assumes risk neutrality for healthy years runs into the additional problem of not being empirically realistic, because risk aversion is the common empirical finding.

The calculation mistake in Example 2 of Mehrez and Gafni.¹ Let us, finally, consider Example 2 of Mehrez and Gafni,^{1(end of p.146)} which has given rise to much discussion. Mehrez and Gafni consider there the assumption that true preferences deviate from the assumptions of the QALY model (the QALY model includes, in this discussion, the assumption of risk neutrality). Obviously, this assumption is logically equivalent to the existence of examples where one treatment is preferred according to true preferences, and the other treatment according to the QALY model. For clarification, Mehrez and Gafni suggest their Example 2 as a hypothetical numerical example to illustrate such a “preference reversal,” where the true preferences differ from the QALY preferences. Unfortunately, a calculation mistake slips in in their example.

The mistake is illustrated in figure 1. Mehrez and Gafni claim that the QALY model would derive from the indifference in figure 1c that the quality weight for health state Q_1 be 0.65. A comparison with figures 1a and 1b shows what the mistake is? figure 1c is an unfortunate mix of figures 1a and 1b. “The QALY model does conclude both from the indifference in figure 1a and from the indifference in figure 1b that the quality weight for Q_1 is 0.65. Notice here that in both figures, the same life duration appears to the left and the right sides of the indifference sign (10 yr in figure 1a and 5 yr in figure 1b), as it should be. In figure 1c, however, the left life duration of 5 years as in figure 1b has been combined with the right life duration of 10 years as in figure 1a. The indifference in figure 1c contradicts the indifferences in figures 1a and 1b, and by the QALY model contradicts a quality weight of 0.65 for Q_1 as claimed

by Mehrez and Gafni. § The calculation mistake is, by itself, not a very serious issue, of course. What must be regretted, however, is that Gafni et al.^{6,9} do not acknowledge their calculation mistake.

Do HYE “fully represent individual’s preferences?” Theorem 1 above demonstrates that HYE do not represent risk attitudes, such as the commonly found risk aversion. Therefore, they cannot “fully represent individuals’ preferences,” contrary to claims to the contrary.’

Another, related, misunderstanding may concern the following point. One can generalize HYE in such a way as to incorporate risk attitudes, and, in effect, the whole preference structure of a person. We call these values “extended” HYE. ¶ The procedure is as follows. For each gamble, the extended HYE is defined as the number of **healthy years** equivalent to the **gamble**. So HYE are then defined over all gambles, rather than only over sure outcomes. Then one gamble is preferred over another if and only if it has a higher extended HYE value. For extended HYE, the claim of full representation of individuals’ preferences (including risky choices) is justified. Here lies, however, also the disadvantage of extended HYE. Recording the information concerning extended HYE comes down to recording the whole preference ordering. A model that describes reality by simply repeating every aspect of reality explicitly, without providing any convenient summary, is not useful. Therefore, extended HYE cannot replace utilities in decision analysis.

HYE can be useful in communication. If one describes to non-specialist lay-persons, such as most patients, the value of a medical treatment in terms of its expected utility value, the laypersons will not understand. A lay-person knows the meaning of 9.5 years in perfect health, but does not know the meaning of 0.97 utility. Expected utility is a theoretical notion that has no direct empirical meaning. It is therefore easier to express the value of a medical treatment in terms of the equivalent number of healthy years, i.e., the HYE. For this purpose, the HYE are useful.”

Notice here that, for optimal decision making, the decision analyst should retain the full information of the von Neumann-Morgenstern (vNM) utility functions. It is easy to transform utilities into HYE, as explained above. However, as implied by Theorem 1 above, transforming HYE into vNM utilities

‡Also, some critics of HYE seem to assume that the HYE approach adopts other assumptions from TTO calculations, such as valuation of health state Q through the quotient X/Y [from an indifference $(Q, Y) \sim (H, X)$], or other assumptions from QALY calculations. To my knowledge, such assumptions have never been made for HYE.

§The original text of the example could be interpreted in two ways, leading to two different preferences that would both lead to a calculation mistake. Gafni et al.^{6,9} state that they intend the example as it is depicted in figure 1c.

¶References 5 and 17 use the expression “certainty equivalent number of (healthy) years.”

is not possible: If one has recorded all the HYE values for all health profiles, then one does not possess information about risk attitude and cannot retrace the associated vNM utilities. This analysis shows that storing information in terms of HYE, instead of in terms of vNM utilities, is an unfortunate choice by which the information about risk attitudes is lost. This is contrary to claims by Gafni et al.^{1,6}

Empirical Performance of HYE: Dropping the Theoretical Assumption of Expected Utility

The main body of the paper presents, in line with the discussions in the field, a theoretical analysis of HYE, based on rationality principles. That way we find that HYE are identical to TTO, and that this identity is based mainly on the transitivity condition. However, eliciting utilities from patients is a descriptive activity, and rationality principles are violated descriptively. For instance, the probability-equivalent version of a standard gamble and the certainty-equivalent version should give identical utility functions according to expected utility, but in reality it is generally found that the probability-equivalent method gives higher values.²⁸ This section addresses such descriptive issues.

On the basis of phenomena such as those described above, it can be expected that in empirical investigations differences will be found between the information elicited by direct TTO questions and by the two-stage procedure advocated by Gafni et al. This does not yet justify their approach. For that, an empirical theory and explanations should be provided for the deviations from rationality that occur, and it should be explained why the two-stage method provides better protection against such deviations. Such theory and explanations have not yet been offered by Gafni et al., who reiterate that their ideas are intended to be based on expected utility theory. For instance, Mehrez and Gafni^{1, p.145} state explicitly that their ideas are not motivated by the empirical deviations from rationality. They claim^{3, p.168} that one has to follow the measurement procedures that coincide with expected utility if one accepts expected utility as the normative standard. (This is contrary to the view that one should correct elicitation for systematic descriptive deviations from the normative theory.)

At any rate, I expect that the two-stage approach of Gafni et al. will not provide better empirical results than TTO questions, but, on the contrary, will only generate additional errors and distortion because of its more complicated nature. (Such an expectation has been expressed before.[?] There is some empirical evidence available in the literature

that supports that conjecture. It has been found that for the probability-equivalent method adopted in the first stage of the two-stage measurement procedure proposed by Mehrez and Gafni, response-mode effects lead to overestimation of the probability values (enhancing risk aversion)^{29,30} For the certainty-equivalent questions asked in the second stage, response-mode effects lead to overestimation of the certainty equivalents (reducing risk aversion).^{29,30} These two effects both increase the observed HYE value. Thus, the deviations from transitivity, i.e., the factors to generate differences between the stages of the two-stage approach of Gafni et al., do not cancel but reinforce each other. This predicts that HYE observed by means of the two-stage procedure of Gafni et al. lead to strong overestimations, stronger than those obtained using TTO observations.

Conclusion

Under expected utility, as assumed by Mehrez, Gafni, and their colleagues:

1. The two-stage approach to elicit HYE is a round-about way for measuring TTO values.
2. The two-stage approach does not elicit risk attitude; Gafni et al. thought otherwise because they did not understand that expected-utility theory also governs riskless choices and that in the value-utility theory, riskless value functions and risky utility functions order outcomes in the same manner.
3. The greater generality of HYE as compared with QALYs is not a virtue per se ("proved" by Gafni et al. through their "counterexamples"), but must be weighed against the loss of tractability.
4. If risk neutrality is assumed for healthy life years, then this assumption deviates from empirical reality.
5. If no risk neutrality is assumed for healthy life years, then the HYE model is not new, but is simply the general expected-utility model.

If expected utility is not assumed, and the two-stage approach to HYE is tested empirically, then:

6. The two stages enhance errors and lead to overestimations of the HYE values that are stronger than those found with direct TTO questions.

Finally,

7. For communication purposes, HYE seem to be a useful tool.