



Cost-Effectiveness Risk-Aversion Curves: Comparison of Risk-Adjusted Performance Measures and Expected-Utility Approaches

Elamin H. Elbasha^{1,2}

Accepted: 29 November 2021 / Published online: 9 February 2022

© Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, N.J., U.S.A., under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

Accounting for risk attitudes in medical decision making under uncertainty has attracted little research. A recent proposal recommended using the results of a cost-effectiveness analysis to construct a cost-effectiveness risk-aversion curve (CERAC) to inform risk-averse decision makers choosing among healthcare programs with uncertain costs and effects. The CERAC is based on a risk-adjusted performance measure widely used in financial economics called the Sortino ratio. This paper evaluates the CERAC based on the Sortino ratio, derives its various properties, discusses the implications of using it to inform decision making under uncertainty, and compares it with the expected-utility approach. Analytic formulae for the CERAC, relating it to the means and standard deviations of costs and effects of a healthcare program, are derived for both approaches. Compared with the expected-utility approach, the CERAC based on the Sortino ratio implicitly assumes that the decision maker is highly risk averse.

Key Points for Decision Makers

A few approaches have been suggested for accounting for risk attitudes in medical decision making under uncertainty.

This paper evaluates the use of the Sortino-ratio-based risk-adjustment measure to inform decision making under uncertainty.

Compared with the theoretically grounded expected-utility approach, the Sortino-ratio-based risk-adjustment measure implicitly assumes that the decision maker is highly risk averse.

If the level of risk aversion of the decision maker is substantially lower than what is implicitly assumed, use of the Sortino-ratio-based risk-adjustment measure will result in recommending sub-optimal healthcare programs.

1 Introduction

The widely used framework of cost-effectiveness analysis (CEA) assumes that decision makers are neutral toward risk in health and medicine [1]. Accounting for risk attitudes in medical decision making under uncertainty may be important [2]. Recognizing this, a few approaches have been suggested, some based on theoretical grounds, while others can be considered ad hoc recommendations. For example, a recommendation was made to the Joint Committee on Vaccination and Immunisation, an independent expert committee that advises the government of the United Kingdom on immunization programs, to require an immunization program to be deemed cost-effective only if 90% of scenarios in a Monte Carlo simulation fall below a £25,000 per quality-adjusted life-year (QALY) threshold [3]. Such somewhat arbitrary and ad hoc approaches can be problematic, therefore risking adoption of suboptimal healthcare programs [4].

Recently, Sendi [5] introduced the cost-effectiveness risk-aversion curve (CERAC) to inform choices under uncertainty for risk-averse decision makers and policy makers. Sendi [5] proposed using the results of a CEA to construct the CERAC based on a risk-adjusted performance measure widely used in financial economics called the Sortino ratio [6].

The purpose of this paper is to evaluate the Sortino-ratio-based CERAC as a risk-adjustment measure, investigate its

✉ Elamin H. Elbasha
elamin_elbasha@merck.com

¹ Merck & Co., Inc., Kenilworth, NJ, USA

² WP37A-150, PO Box 1000, West Point, PA 19486, USA

various properties, discuss the implications of using it to inform decision making under uncertainty, and compare it with the expected-utility approach. This will be achieved by deriving analytic formulae for both the Sortino-ratio-based CERAC and the expected-utility-based CERAC, relating the CERAC to the moments (e.g., means and standard deviations) of the distribution of costs and effects of a healthcare program. The paper also discusses the level of risk aversion implicitly assumed in the Sortino-ratio-based CERAC.

2 CERAC Based on the Sortino Ratio

The financial economics literature includes several risk-adjusted performance measures that are frequently used to evaluate investment opportunities. Sendi [5] discussed the Sharpe ratio [7, 8], considered to be the most popular performance ratio, and proposed its derivative, the Sortino ratio, to construct the CERAC. The net benefit-to-risk ratio S_{NMB} , analogous to the Sortino ratio, is calculated as the average net monetary benefit (NMB) of a program adjusted by its associated downside risk (defined as the risk of achieving an NMB below the expected NMB). Given the joint distribution of costs (denoted by C) and effects (denoted by E), the NMB of a given program is defined as

$$\text{NMB} = \lambda E - C, \quad (1)$$

where λ is the decision maker's maximum willingness to pay (WTP) per health outcome E (e.g., QALYs). Thus, S_{NMB} is calculated as

$$S_{\text{NMB}} = \frac{\mu_{\text{NMB}}}{\text{DD}_{\text{NMB}}}, \quad (2)$$

where μ_{NMB} is the expected NMB of a program and DD_{NMB} is the root mean square of the downside deviations of the program's NMB realizations from the expected NMB. The expected value of NMB of a program (Eq. 1) is given by

$$\mu_{\text{NMB}} = \lambda \mu_E - \mu_C, \quad (3)$$

where μ_E denotes the expected effects and μ_C denotes the expected costs of the program. The formula for DD_{NMB} of a continuous random variable NMB with probability density function $f(x)$ is defined as:

$$\text{DD}_{\text{NMB}} = \sqrt{\int_{-\infty}^{\infty} (x - \mu_{\text{NMB}})^2 g(x) f(x) dx}, \quad (4)$$

where

$$g(x) = 1 \quad \text{if } x < \mu_{\text{NMB}}$$

$$g(x) = 0 \quad \text{if } x \geq \mu_{\text{NMB}}.$$

It should be mentioned that the Sharpe ratio penalizes all deviations, not just downside deviations only, as in the Sortino ratio. Thus, in deriving the Sharpe ratio, the $g(x)$ function in the above formula becomes $g(x) = 1$ for all x . For both risk-adjustment measures, it is assumed that a risk-averse decision maker prefers a program with a higher net benefit-to-risk ratio.

Although the approach does not require making distributional assumptions, Sendi [5] assumed that both costs and effects are normally distributed, with standard deviation of costs denoted by σ_C , standard deviation of effects denoted by σ_E , and the correlation between the costs and effects of each program denoted by ρ . Thus, it follows from the bivariate normal distribution of costs and effects $\mathcal{N}[(\mu_E, \mu_C), (\sigma_E, \sigma_C), \rho]$ that the NMB of a program is also normally distributed, with the mean given by Eq. (3) and variance var_{NMB} by

$$\text{var}_{\text{NMB}} = \lambda^2 \sigma_E^2 - 2\lambda\rho\sigma_E\sigma_C + \sigma_C^2. \quad (5)$$

Because the normal distribution is symmetric around the mean, the total of upside deviations and the total of downside deviations are equal. Thus, considering only downside deviations from the mean requires only dividing the variance by 2 to obtain DD_{NMB} . Therefore, the root mean square of the downside deviations is

$$\text{DD}_{\text{NMB}} = \sqrt{\frac{\text{var}_{\text{NMB}}}{2}} = \sqrt{\frac{(\lambda^2 \sigma_E^2 - 2\lambda\rho\sigma_E\sigma_C + \sigma_C^2)}{2}}. \quad (6)$$

Combining Eqs. (2), (3), and (6), the net benefit-to-risk ratio S_{NMB} based on the Sortino ratio under normally distributed costs and effects is given by

$$S_{\text{NMB}} = \frac{\mu_{\text{NMB}}}{\sqrt{\frac{\text{var}_{\text{NMB}}}{2}}} = \frac{\sqrt{2}(\lambda\mu_E - \mu_C)}{\sqrt{\lambda^2 \sigma_E^2 - 2\lambda\rho\sigma_E\sigma_C + \sigma_C^2}}. \quad (7)$$

For a normally distributed NMB, the benefit-to-risk ratio based on the Sharpe ratio is

$$\text{Sharpe}_{\text{NMB}} = \frac{\mu_{\text{NMB}}}{\sqrt{\text{var}_{\text{NMB}}}} = \frac{S_{\text{NMB}}}{\sqrt{2}}. \quad (8)$$

It is clear from Eq. (8) that, under the assumption of a normal distribution of NMB, the Sharpe ratio is proportional to the Sortino ratio. Consequently, use of either the Sharpe or the Sortino ratio will result in the same ranking of healthcare programs when the distribution of NMB is symmetric.

As an illustration of the use of the net benefit-to-risk ratio S_{NMB} , Sendi [5] provided three examples; two of

them (example 1 and 2) are shown in Table 1. In example 1, in Table 2, assuming $\lambda = \$50,000/\text{QALY}$, program B has a greater expected NMB (i.e., μ_{NMB} of \$850,000 vs \$700,000) and hence is preferred by a risk-neutral decision maker to program A. However, with greater uncertainties in both costs and effects, program B has a lower net benefit-to-risk ratio (i.e., S_{NMB} 3.1 vs 10.1) and should not be preferred by a risk-averse decision maker. The decision is reversed in example 2. Program D is now preferred (i.e., S_{NMB} 12.3 vs 2.5), because program D has less uncertainty in both costs and effects compared with program C.

2.1 Properties of CERAC Based on the Sortino Ratio

It is important to understand the mathematical properties of the CERAC given by Eq. (7). It should be noted that all these properties are derived under the assumption that both costs and effects are normally distributed. As a function of WTP λ , the CERAC:

1. Cuts the x -axis where the expected NMB is zero (i.e., $S_{\text{NMB}} = 0$ when $\lambda = \mu_{\text{C}}/\mu_{\text{E}}$).
2. Asymptotically approaches the ratio of the expected effects to the square root of half the variance of effects (i.e., $\lim_{\lambda \rightarrow \infty} S_{\text{NMB}} = \mu_{\text{E}}/\sqrt{\sigma_{\text{E}}^2/2}$)
3. Is not always monotonically increasing in λ (Fig. 1 shows cases of both increasing and decreasing CERACs).

Although he used simulation rather than an analytical approach and only a limited number of examples, Sendi [5] correctly described properties (1) and (2). Also, in all the three examples provided by Sendi [5], the CERAC is depicted to be monotonically increasing as λ increases. As shown in Fig. 1, using inputs $\mu_{\text{E}} = 10$, $\mu_{\text{C}} = 60,000$, $\sigma_{\text{E}} = 1.1$, $\sigma_{\text{C}} = 20,000$, and $\rho = 0.9$, the CERAC of a given program is not always monotonically increasing in λ . As a function of λ , the net benefit-to-risk ratio of the program increases initially, reaches a peak, and declines thereafter, reaching its asymptotic value from above.

Table 1 Costs and effects inputs of nine hypothetical healthcare programs

Example	Program	μ_{C} (\$)	σ_{C} (\$)	μ_{E} (QALY)	σ_{E} (QALY)	ρ
Example 1	A	50,000	5000	15	2	0.4
	B	150,000	20,000	20	8	0.4
Example 2	C	50,000	20,000	15	8	0.4
	D	150,000	5000	20	2	0.4
Example 3	A	50,000	5000	15	2	0.4
	B'	150,000	20,000	59	8	0.4
	B''	150,000	20,000	17	8	0.4
Example 4	C'	50,000	20000	69	8	0.4
	D	150,000	5000	20	2	0.4

μ_{C} denotes mean costs, σ_{C} denotes standard deviation of costs, μ_{E} denotes mean effects, σ_{E} denotes standard deviation of effects, and ρ denotes correlation between costs and effects of each program

Joint normal distributions for costs and effects are assumed

QALY quality-adjusted life-year

Table 2 ICER, NMB, and risk-adjusted outcomes of nine hypothetical healthcare programs

Example	Program	ICER (\$/QALY)	μ_{NMB} (\$)	σ_{NMB} (\$)	S_{NMB}	CE_{NMB} (\$)
Example 1	A	–	700,000	98,107	10.1	651,875
	B	20,000	850,000	392,428	3.1	80,000
Example 2	C	–	700,000	392,428	2.5	(70,000)
	D	20,000	850,000	98,107	12.3	801,875
Example 3	A	–	700,000	98,107	10.1	651,875
	B'	2273	2,800,000	392,428	10.1	2,030,000
	B''	50,000	700,000	392,428	2.5	(70,000)
Example 4	C'	–	3,400,000	392,428	12.3	2,630,000
	D	Dominated	850,000	98,107	12.3	801,875

μ_{NMB} denotes mean NMB, σ_{NMB} denotes standard deviation of NMB, CE_{NMB} denotes the certainty equivalent, and S_{NMB} denotes the net benefit-to-risk ratio. To calculate values of CE_{NMB} and S_{NMB} , it is assumed that the value of the willingness to pay λ is \$50,000 and the absolute risk-aversion parameter r is 0.00001

ICER incremental cost-effectiveness ratio, NMB net monetary benefit, QALY quality-adjusted life-year

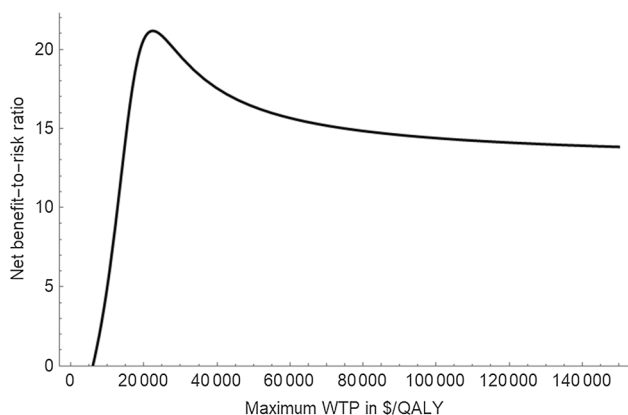


Fig. 1 CERAC (parameters values: $\mu_E = 10$, $\mu_C = 60,000$, $\sigma_E = 1.1$, $\sigma_C = 20,000$, $\rho = 0.9$). The net-benefit ratio increases initially with WTP, reaches a peak, and declines thereafter, reaching its asymptotic value from above. μ_C denotes mean costs, σ_C denotes standard deviation of costs, μ_E denotes mean effects, σ_E denotes standard deviation of effects, and ρ denotes correlation between costs and effects of each program. Joint normal distributions for costs and effects are assumed. CERAC cost-effectiveness risk-aversion curve, WTP willingness to pay, QALY quality-adjusted life-year

In general, the shape of the CERAC, including non-monotonicity and the position of the peak, is ambiguous and is determined by the interplay of many factors. For example, when the correlation between costs and effects approaches zero, the CERAC is always monotonically increasing. However, when there is correlation between costs and effects, CERAC can be monotonically increasing or has an inverted-U shape.

For a given value of WTP λ , the net benefit-to-risk ratio increases with higher mean values of effects μ_E and lower values of mean costs μ_C , and higher values of the correlation between costs and effects ρ for all positive values of NMB. The positive association between net benefit-to-risk ratio and ρ follows from the inverse relationship between the variance of NMB and the correlation between costs and effects ρ . The net benefit-to-risk ratio can increase or decrease with higher values of the standard deviation of costs σ_C or effects σ_E . As a function of the overall mean and variance of NMB, the net-benefit ratio increases with increases in the mean and decreases, albeit at a lower rate, with lower variance of NMB. Mathematically speaking, the net benefit-to-risk ratio is homogenous of degree zero in the mean and standard deviation of NMB. That is, the net benefit-to-risk ratio does not change if both the mean and standard deviation of NMB change at the same rate. For example, doubling the values of the mean and standard deviation of costs and effects leaves the ratio unchanged. Sendi [5] did not describe this very important property of the net benefit-to-risk ratio.

The last property of the net benefit-to-risk ratio is more general according to Eq. (2): the ratio does not change if

both the mean and the root mean square of the downside deviation of NMB change at the same rate. In other words, proportional changes in the overall mean and the root mean square of the downside deviation of NMB are valued equivalently. This raises the following question: can CERAC correctly account for the trade-off between risk and expected returns? Because the net benefit-to-risk ratio as proposed by Sendi [5] is not based on theoretical grounds, it can be considered somewhat arbitrary. For example, there is no justification for implicitly assuming a fixed risk aversion that cannot be varied across decision makers. More importantly, what is the risk attitude implicit in the use of the Sortino-based net benefit-to-risk ratio performance measures?

Example 3 illustrates how much expected value a risk-averse decision maker will be willing to forgo in order to avoid risk (Table 1). The inputs in example 3 are the same as those in example 1, except that the mean effect of program B' is now almost close to three times the original mean (20 vs 59 QALYs). Example 3 also includes a new scenario B'' where the mean effect of program B'' is 17 QALYs. Recall that in example 1 program B is preferred by a risk-neutral decision maker whereas program A is preferred by a risk-averse decision maker. In example 3, the incremental cost-effectiveness ratio (ICER) of program B' compared with program A is \$2273/QALY (Table 2). When the mean effect of program B'' is 17 QALYs, the ICER of B'' compared with A is equal to \$50,000/QALY, and a risk-neutral decision maker is indifferent between program A and B''. At mean effect of program B' of 59 QALYs, both programs have a net benefit-to-risk ratio of 10.1, and a risk-averse decision maker is indifferent between program A and B'. This implies that, compared with a risk-neutral decision maker, a risk-averse decision maker will only be willing to accept an additional 42 (= 59 - 17) QALYs per person (i.e., equivalent to an incremental NMB of \$2,100,000) in order to become indifferent between program A and the riskier program B'.

Example 4 is the same as example 2, except that the mean effect of program A' is now 69 instead of 15 QALYs (Table 1). Recall that in example 2, program D is preferred over program C by either a risk-neutral or a risk-averse decision maker. However, assuming $\lambda = \$50,000/\text{QALY}$, example 4 illustrates that program C' is preferred over program D by a risk-averse decision maker but is no longer preferred by a risk-neutral decision maker as long as the mean effects of program C' is greater than 18 QALYs (Table 2). In fact, according to the net benefit-to-risk ratio approach, even though it is strongly being dominated by program C', program D is always preferred over program C' by a risk-averse decision maker as long as program C' offers mean effects of less than 69 QALYs. In other words, compared with a risk-neutral decision maker (indifferent between the two programs if the mean effect is 18 QALYs), a risk-averse decision maker will be willing to

accept at least 51 (= 69 – 18) additional QALYs (equivalent to an incremental NMB of \$2,550,000) in order to prefer program C’ over the less risky program D.

Both example 3 and 4 in Table 1 illustrate a very important property of the Sortino-ratio-based CERAC: the preference of a risk-averse decision maker using the Sortino-ratio-based CERAC implies a strong risk–return trade-off. For example, suppose a hypothetical program E has double the standard deviation (i.e., quadruple the variance) of NMB compared with another program F. The two programs are valued equally by a risk-averse decision maker using the Sortino-ratio-based CERAC only if the mean NMB of program E is also double that of program F.

In a more recent paper, Sendi et al. [9] claim that the downside deviation can be modified to reflect different levels of risk aversion. The authors suggest that this can be done by defining a different minimally acceptable return (MAR) for the downside deviation of NMB. The authors were not clear on how to select an MAR, nor on what is an appropriate MAR, but discussed a few options—all seem arbitrary. Focusing on their first alternative choice of defining the downside deviation relative to the 25% percentile of the NMB distribution, denoted as $\eta_{25\text{NMB}}$, instead of the mean, the denominator of the ratio, Eq. (4), becomes

$$DD25_{\text{NMB}} = \sqrt{\int_{-\infty}^{\infty} (x - \eta_{25\text{NMB}})^2 g(x)f(x)dx}, \tag{9}$$

where

$$g(x) = 1 \quad \text{if } x < \eta_{25\text{NMB}}$$

$$g(x) = 0 \quad \text{if } x \geq \eta_{25\text{NMB}}.$$

Applying this to the inputs of example 3 resulted in higher ratios because the downside deviation is lower now (\$37,920.3 for program A and \$151,681 for program B’). Most importantly, both programs have the same net benefit-to-risk ratio of 18.5, signifying that a risk-averse decision maker is still indifferent between program A and B’. In other words, the level of risk-aversion did not change whether MAR was defined as the mean or the 25% percentile of the NMB. This example suggests that, regardless of the choice of MAR, using the Sortino-ratio-based CERAC implicitly assumes that the decision maker is highly risk averse.

The result that changing MAR does not modify the degree of risk aversion can be shown to be more general than the inputs of example 3 by deriving a general formula for the Sortino ratio $S\alpha_{\text{NMB}}$ of a normally distributed NMB for any percentile α (where $0 < \alpha < 1$):

$$S\alpha_{\text{NMB}} = f(\alpha) \frac{\mu_{\text{NMB}}}{\sqrt{\text{var}_{\text{NMB}}}}, \tag{10}$$

where μ_{NMB} and var_{NMB} are as defined before by Eqs. (3) and (5), respectively. The well-defined function $f(\alpha)$ is independent of the mean and variance of the NMB distribution. When MAR is given by the mean ($\alpha = 50\%$), $f(\alpha) = \sqrt{2}$. The value of $f(25\%) = 1.4142$ when MAR is equal to the 25% percentile of the NMB. It is clear from Eq. 10 that the choice among healthcare programs is the same regardless of the value of α . Consequently, whether MAR is defined as the mean or any percentile of the NMB, use of the Sortino-ratio-based CERAC will result in the same ranking of healthcare programs.

3 CERAC Based on Expected-Utility Approach and Certainty Equivalence

The health economics literature includes a few attempts at incorporating decision makers’ risk preferences when making decisions under uncertainty using the standard expected-utility maximization methods [10–13]. For example, Elbasha [13] suggested using the exponential utility–moment-generating function (mgf) approach for incorporating risk preferences and the trade-off between risk and returns to investment in health and medicine. This approach assumes that the utility function of each individual is represented by an exponential function, a convenient functional form allowing the expected value of utility function to be expressed in terms of only the mgf of the random variables. From welfare economics, a utilitarian welfare function assumes equal individual weights (since individuals are assumed to be *ex ante* identical) and linearly aggregates individual exponential utility functions. The exponential, also known as the constant absolute risk aversion (CARA), utility function over NMB is given by

$$U(\text{NMB}) = -e^{-r\text{NMB}}, \tag{11}$$

where r is the absolute risk-aversion parameter. Applying expectation, the expected utility is

$$\mathbb{E}[U(\text{NMB})] = -\mathbb{E}[e^{-r\text{NMB}}] = -M(-r), \tag{12}$$

where \mathbb{E} is the expectation operator and $M(-r)$ is the mgf of the distribution of the random variable NMB. The mgf of the normal distribution of NMB $M(-r)$ is

$$M(-r) = e^{-r\left(\mu_{\text{NMB}} - r \frac{\text{var}_{\text{NMB}}}{2}\right)}. \tag{13}$$

It is well-known that maximizing expected utility is the same as maximizing the certainty equivalent (CE) version of the expected utility [13]. The CE is defined as the fixed

sure amount that one is willing to accept to be indifferent to a risky position. Thus, the CE of the NMB is defined as

$$U(\text{CE}) = \mathbb{E}[U(\text{NMB})].$$

Using Eqs. (11) and (12), the CE of the expected negative exponential utility function satisfies the following equation:

$$-e^{-r\text{CE}} = -M(-r),$$

which can be solved for CE as

$$\text{CE} = -\frac{\ln M(-r)}{r}. \quad (14)$$

Substituting Eq. (13) into Eq. (14), the CE of a normally distributed NMB simplifies to the well-known mean-variance CE

$$\text{CE}_{\text{NMB}} = \mu_{\text{NMB}} - \frac{r}{2} \text{var}_{\text{NMB}}. \quad (15)$$

Utilizing Eqs. (3), (5), and (15), the CE can be expressed in terms of the parameters of the bivariate normal distribution of costs and effects as

$$\text{CE}_{\text{NMB}} = \lambda\mu_E - \mu_C - \frac{r}{2} \left(\lambda^2 \sigma_E^2 - 2\lambda\rho\sigma_E\sigma_C + \sigma_C^2 \right). \quad (16)$$

Thus, if costs and effects follow a joint normal distribution, assuming that the value of the risk-aversion parameter r is known, Eq. (16) can be used to calculate the CE of each program. A risk-averse decision maker prefers a program with a larger CE. Hence, a risk-averse decision maker should choose the program with the highest CE value.

In example 1 in Table 2, assuming $\lambda = \$50,000/\text{QALY}$ and $r = 0.00001$, program A is preferred by a risk-averse decision maker because it has a higher CE value (i.e., CE_{NMB} of \$651,875 vs \$80,000). As with the net benefit-to-risk ratio approach, a risk-averse decision maker prefers program A in example 1 in Table 1 according to the CE approach as long as the value of r is greater than 0.0000021. However, a less risk-averse decision maker (e.g., $r < 0.0000021$) would prefer program B to program A. By varying the values of λ , the CERACs based on the CE can also be constructed. For example, the CE of program B increases initially with WTP, reaches a peak, and declines thereafter (Fig. 2).

In example 2, according to the CE approach, program D is preferred over program C (Table 2). In fact, any risk-averse decision maker (i.e., r is non-negative) would always prefer program D over program C as long as $\lambda \geq \$20,000/\text{QALY}$. Similarly, according to the CE approach, program B', in example 3, is preferred over program A (Table 2). Moreover, for a risk-averse decision maker to be indifferent between program A and program B' in example 3 according to the CE approach, the risk-aversion parameter r must be 0.0000291 (Fig. 4). Recall that the level of risk in example

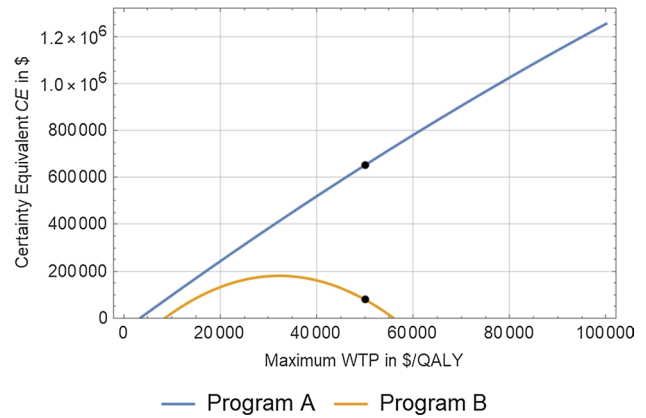


Fig. 2 CERAC based on the CE (example 1 in Table 1). The CE increases initially with WTP, reaches a peak, and declines thereafter. For example, when WTP is \$50,000/QALY and r is 0.00001, program A has a CE of \$651,875 (black point on the blue curve) and program B has a CE of \$80,000 (black point on the orange curve). CE certainty equivalent, CERAC cost-effectiveness risk-aversion curve, r absolute risk-aversion parameter, WTP willingness to pay, QALY quality-adjusted life-year.

1 and example 3 are the same. Thus, comparing the risk-aversion parameters (0.0000291 vs 0.0000021), by requiring higher expected return to be willing to be indifferent between program A and program B' in example 3, the decision maker following the net-benefit ratio approach is 14 times more risk averse compared with the risk-averse decision maker in example 1.

To better understand the trade-off between risk and return as the WTP varies and the role of the decision maker's attitude to risk, it is instructive to draw a cost-effectiveness risk-aversion map (CERAM) that depicts combinations of the risk-aversion parameter r and WTP λ for which the risk-averse decision maker is indifferent between two programs and areas where the program is being preferred over another. In example 1 (Table 1), the blue curve in Fig. 3 depicts combinations of λ and r where program A has the same CE as program B, and the blue area represents combinations where program A has larger CE than program B. Similarly, the orange area represents combinations where program A has lower CE than program B. For example, when WTP is \$50,000/QALY and r is 0.00001, a risk-averse decision maker would prefer program A over program B. However, when WTP is \$50,000/QALY, a less risk-averse decision maker (e.g., $r < 0.0000021$) would prefer program B over the less risky program A.

A similar CERAM can be constructed using the data from example 3, in Table 1 (Fig. 4). With WTP = \$50,000/QALY and $r = 0.00001$, a clear choice for a risk-averse decision maker is the riskier program B' with higher expected return. Program B' remains the optimal choice unless the level of risk aversion for the decision maker is higher

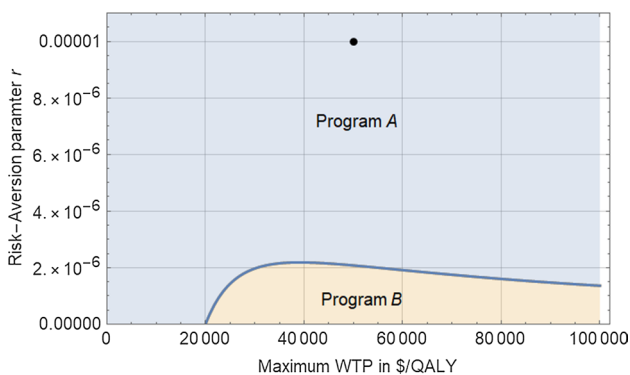


Fig. 3 Combinations of risk-aversion parameter r and WTP that determine the optimal program based on CE (example 1 in Table 1). Blue area represents combinations where program A has a larger CE than program B. Combinations where program B is preferred over program A are given by the orange area. For example, when WTP is \$50,000/QALY and r is 0.00001 (black point), a risk-averse decision maker would prefer program A over program B. *CE* certainty equivalence, r absolute risk-aversion parameter, *WTP* willingness to pay, *QALY* quality-adjusted life-year

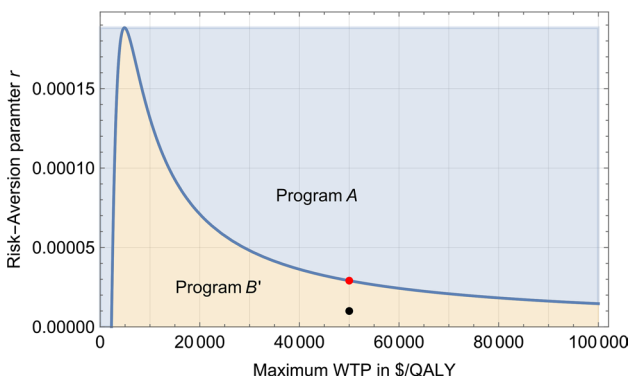


Fig. 4 Combinations of risk-aversion parameter r and WTP that determine the optimal program based on CE (example 3 in Table 1). Blue area represents combinations where program A has a larger CE than program B'. Combinations where program B' is preferred over program A are given by the orange area. For example, when WTP is \$50,000/QALY and r is 0.00001 (black point), a risk-averse decision maker would prefer program B' over program A. When WTP is \$50,000/QALY and r is 0.0000291 (red point), a risk-averse decision maker would be indifferent between program B' and program A. *CE* certainty equivalence, r absolute risk-aversion parameter, *WTP* willingness to pay, *QALY* quality-adjusted life-year

than 0.0000291. If the $WTP = \$50,000/QALY$ and $r > 0.0000291$, a risk-averse decision maker would prefer the less risky program A.

3.1 Properties of CERAC Based on the Certainty Equivalence

It is important to understand the mathematical properties of the CERAC given by Eq. (14). All other things being

equal, the CE increases with mean levels of effects and higher values of the correlation between costs and effects ρ and decreases with mean levels of costs, and risk aversion. The CE can increase or decrease with higher values of the standard deviation of costs σ_C or effects σ_E . As a function of the overall mean and variance of NMB, the CE increases with increases in the mean and decreases, albeit at a lower rate, with lower variance of NMB. CE is an inverted-U shape as function of λ , being negative for low values of λ , increases as λ increases until it reaches a maximum (when $\lambda = \frac{\mu_E + r\rho\sigma_C\sigma_E}{r\sigma_E^2}$), and decreases thereafter.

3.2 An Alternative Utility Function

The CE-based CERAC assuming an exponential utility function affords the opportunity to derive analytic results for any distribution of costs and effects since most known distributions have mgf. Nonetheless, the CE-based CERAC can be constructed using other utility functions. Some of these functional forms, such as the quadratic utility function, also allow derivation of analytic results. However, the quadratic utility function has some theoretically unsatisfactory properties (it decreases as NMB increases over some ranges, and it exhibits increasing absolute risk aversion). Another popular functional form, considered to be the most widely used for empirical analysis [14], is the constant relative risk aversion (CRRA) utility function [15]

$$U(NMB) = \frac{NMB^{1-\gamma} - 1}{1 - \gamma}, \tag{15}$$

where γ is the coefficient of CRRA. The CE derived from CRRA is given by

$$CE_{NMB} = (1 + (1 - \gamma)E[U(NMB)])^{\frac{1}{1-\gamma}}. \tag{16}$$

If NMB is lognormally distributed, with $\log(NMB)$ being normally distributed with mean m and standard deviation s , an analytic formula for CE can be derived as

$$CE_{NMB} = e^{m - \frac{1}{2}(\gamma-1)s^2}. \tag{17}$$

As with CARA, a risk-averse decision maker should choose the program with the highest CE_{NMB} derived using a CRRA utility function. To illustrate, the parameters in example 1, in Table 1, assuming $\lambda = \$50,000/QALY$, result in a mean NMB of \$700,000 and \$850,000 for program A and B, respectively (Table 2). The respective standard deviation is 98,107.1 and 392,428. These values are related to the parameters of the lognormal distribution μ_{NMB} and σ_{NMB} according to

$$m = \ln \left(\frac{\mu_{\text{NMB}}^2}{\sqrt{\mu_{\text{NMB}}^2 + \sigma_{\text{NMB}}^2}} \right),$$

$$s = \sqrt{\ln \left(1 + \frac{\sigma_{\text{NMB}}^2}{\mu_{\text{NMB}}^2} \right)}.$$

This implies values of $(m, s) = (13.449, 0.139)$ for program A and $(m, s) = (13.556, 0.439)$ for program B. Substituting these values into Eq. 17 and assuming a CRRA coefficient of $\gamma = 2.5$, program A has a CE value of \$683,184, whereas program B has a CE value of \$667,615. Therefore, according to the CE approach based on a CRRA utility function, a risk-averse decision maker prefers program A in example 1 in Table 1. In fact, assuming $\lambda = \$50,000/\text{QALY}$, program A is always preferred over program B as long as the value of γ is greater than 2.23467. However, a less risk-averse decision maker (e.g., $\gamma < 2.23467$) would prefer program B over program A.

4 Discussion

Standard CEA frameworks do not consider the role of risk attitudes in informing coverage and reimbursement decisions of new medical technologies under uncertainty. There has been little research that has emphasized the importance of incorporating risk preferences and the trade-off between the mean and the risk of returns to investment in health and medicine [9, 10, 16–19]. This paper presents two approaches for extending CEA to incorporate attitudes toward risk and contrasted them. The first approach applies principles of portfolio theory of how individual investors select among risky financial investments to CEA of healthcare programs. Specifically, the paper analyzes the approach introduced by Sendi [5] and Sendi et al. [9] that uses the net benefits-to-risk ratio derived from the popular Sortino financial risk-adjustment measure. Some of the results regarding the CERAC based on a Sortino ratio depend on the assumption of normality. It is not clear what would be the implications of deviations from this assumption. However, it is expected that major conclusions would still be valid under other distributional assumptions. For example, whether a simulation or an analytical approach is used, or whether a normal distribution is assumed or not, the conclusion that the CERAC based on the Sortino ratio risk-adjustment measure implicitly assumes that the decision maker is highly risk averse remains valid.

The second approach draws from principles of welfare economics and expected utility theory in the evaluation of and selection from risky healthcare programs. The approach is illustrated using two frequently used utility functions:

CARA or CRRA utility functions. The CARA utility function has several advantages, including being a second-order local approximation to any well-behaved utility function. Likewise, the CRRA utility function provides flexible functional forms for modeling risk aversion that can fit the data well.

4.1 A Comparison of the Sortino-Ratio and Expected-Utility Approaches

Each approach has strengths and limitations (Table 3). The major strengths of the Sortino-ratio-based approach to constructing the net benefits-to-risk ratio include [5]:

- S1 There is no need to explicitly specify a preference function over expected return and risk.
- S2 There is no need to make assumptions about the distribution of costs and effects. Instead, it is possible to use empirical NMB distribution obtained from the probabilistic sensitivity analysis (PSA).
- S3 Superior performance measure when the distribution of NMB is skewed and the risk-averse decision maker penalizes expected NMB for only downside deviation without penalizing upside deviation.
- S4 Straightforward to calculate and widely applied in finance.
- S5 Has several desirable qualitative properties (rises with mean and falls with variance of NMB).

The major weaknesses of the Sortino-ratio-based approach include:

- W1 It is an ad hoc approach to characterizing the risk–return trade-off among healthcare programs. Although the risk preferences, implicit in using risk-adjusted performance measures, are not well understood, rankings based on the Sortino ratio may not be consistent with preferences of a decision maker with low or moderate levels of risk aversion [20].
- W2 A CERAC constructed based on a risk-adjusted performance measure like the Sortino ratio includes an implied fixed risk attitude and does not afford the decision maker the opportunity to impose her or his own individual risk attitudes.
- W3 By including only downside deviation, the approach excludes upside deviation, which may be important to a large class of risk-averse decision makers. To illustrate, consider the following simple hypothetical example of a risk-averse decision maker choosing between two risky investment opportunities. The first option can result in a loss of -200 with probability 20% or a gain of 50 with probability 80%. The second option can also result in a loss of -200 with probability 20%, but the gain is

Table 3 Strengths and weaknesses of CERAC based on Sortino-Ratio and expected-utility approaches

Characteristic	CERAC based on Sortino ratio	CERAC based on expected utility	
Strengths	Preferences	No need to specify a preference function	Based on principles of welfare economics Risk preferences vary by type of decision maker
	Assumptions	No need to make assumptions about the distribution of costs and effects Superior performance measure when the distribution of NMB is skewed	No need to make assumptions about the distribution of costs and effects Superior performance measure with any distribution of NMB
	Ease of use	Straightforward and easy to calculate	Straightforward to calculate
	Application	Widely applied in finance and a couple of applications in healthcare	Widely applied in finance and economics, with a few applications in healthcare
	Properties	Has several desirable qualitative properties	Has several desirable qualitative properties
Weakness	Preferences	Ad hoc approach to characterizing the risk–return trade off Implied fixed risk attitude Consistent only with high levels of risk aversion	Need to specify a preference function Requires knowing the degree of risk aversion
	Assumptions	Excludes upside deviation	
	Properties	Dimensionless quantity difficult to interpret a negative net benefits-to-risk ratio	

CERAC cost-effectiveness risk-aversion curve, NMB net monetary benefit

more uncertain: 20 with probability 20% and 60 with probability 60%. Both options have expected values of zero and the same downside deviation of 8000 ($= 0.2 * (-200)^2$). According to the Sortino ratio, the decision maker should be indifferent between the two options because the downside deviations (i.e., uncertainty in losses) are equivalent. However, the first option has an upside deviation (i.e., uncertainty in gains) of 2000, whereas the upside deviation of the second option of 2240 is higher. For a risk-averse decision maker, the first option is clearly better because the upside risk is lower compared with that of the second option. It should be mentioned that this is consistent with how risk aversion is commonly defined in economics and finance as “the tendency of people to prefer outcomes with low uncertainty to those outcomes with high uncertainty, even if the average outcome of the latter is equal to or higher in monetary value than the more certain outcome” [21]. In this example, the average gains are equal ($0.8 * 50 = 0.2 * 20 + 0.6 * 60 = 40$) across the two options, but the uncertainty in gain is lower for the first option compared with that of the second option (2000 vs 2240).

W4 It is difficult to interpret a negative net benefits-to-risk ratio resulting from a negative expected NMB. Also, when the expected NMB is zero, the net benefits-to-risk ratio is zero regardless of the level of risk of a healthcare program.

The expected-utility-based approach shares with the Sortino-ratio-based approach major strengths S2 and S5. In addition, the most important strength of this approach

to decision making under uncertainty is that its theoretical foundation is rooted in traditional welfare economics [12]. However, despite their theoretical appeal (e.g., being based on the principles of welfare economics and expected-utility theory), methods to analyze the results of a CEA under uncertainty and risk aversion have not been widely used. Among the limitations of the CE approach are:

W1. CE typically require the analyst to explicitly characterize preferences over risk and expected return by specifying an explicit utility function. For example, the moment-generating approach assumes an exponential utility function, which implies CARA. This excludes an important class of risk-averse decision makers that have preferences that exhibit decreasing absolute risk aversion or CRRA. Although it is possible to use the approach with a CRRA utility function to derive analytic results, the required assumption of log-normally distributed NMB may not be reasonable given that NMB can be negative for low values of WTP.

W2. CE requires knowing the decision maker’s risk attitude by estimating the degree of risk aversion of the decision maker, which might be difficult to elicit in practice.

Instead of implicitly assuming a fixed degree of risk aversion as in the case of financial risk-adjusted measures, the CE approach explicitly incorporates risk preferences of decision makers. It is true that understanding and eliciting the decision maker’s risk attitude in health is difficult. Despite the availability of methods of estimation of risk attitude coefficients such as the Arrow–Pratt measure of absolute risk-aversion parameter r , there are no accepted estimates

of r [22–24]. One solution suggested in this paper is to use a range of risk-aversion coefficients extending from risk neutrality to moderate risk aversion and depict that against WTP in a two-way sensitivity approach. For example, using the inputs of example 1 (Table 1), the blue shaded area in Fig. 3 shows the combinations of (λ, r) where a risk-averse decision maker would prefer program A over program B. It should be noted that to overcome the difficulty of directly eliciting the decision maker's maximum WTP for per health outcome λ , the cost-effectiveness acceptability curve is constructed by plotting the probability that the joint distribution of incremental costs and effects where the intervention is cost-effective for a range of values of λ . Analogously, the CE-based CERAC can be constructed by plotting pairs of risk-aversion parameter r and λ for which the decision maker is indifferent between two healthcare programs (Figs. 3, 4).

For most practical applications of CEA, there may be no need for choosing an appropriate joint distribution for costs and effects or a distribution of NMB. Instead, the analyst can use the results of the PSA from CEA directly to estimate the mean utility using the equation of the chosen utility function (e.g., Eq. 9 or Eq. 15).

In conclusion, adapting financial risk-adjustment measures such as the Sortino ratio to construct a CERAC can be a helpful tool for highlighting the level of risk to risk-averse decision makers when making choices among healthcare interventions with uncertain costs and effects. However, compared with the expected-utility approach, risk-adjustment measures implicitly assume levels of risk aversion that can be considered high for a class of decision makers. If the level of risk aversion of the decision maker is substantially lower than what is implicitly assumed, it is not recommended to use the Sortino-ratio-based risk-adjustment measure because it will result in sub-optimal decisions. In addition to being theoretically grounded in the principles of welfare economics, the expected-utility approach accommodates wide ranges of risk preferences of decision makers. By presenting a CERAM depicting combinations of risk aversion and WTP where each health program is considered cost-effective, the expected-utility approach has the added advantage of allowing the risk attitudes of the decision maker to be explicitly incorporated.

Declarations

Funding This work was supported by Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA; Merck Sharp and Dohme.

Conflict of interest The author declares no competing interests.

Availability of data and material Not applicable.

Code availability Code is available from the author upon request.

Authors' contributions Not applicable.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

- Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med*. 1977;296(13):716–21.
- Ben-Zion U, Gafni A. Evaluation of public investment in health care. Is the risk irrelevant? *J Health Econ*. 1983;2:161–5.
- Immunisation and High Consequence Infectious Diseases Team, Global and Public Health Group. Cost-effectiveness methodology for vaccination programmes: consultation on the Cost-Effectiveness Methodology for Vaccination Programmes and Procurement (CEMIPP) Report. Department of Health and Social Care; 2018. pp. 1–19. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/707847/cemipp-consultation-document.pdf. Accessed 5 Mar 2021.
- O'Mahony JF, Paulden M. Appraising the cost-effectiveness of vaccines in the UK: insights from the Department of Health Consultation on the revision of methods guidelines. *Vaccine*. 2019;37(21):2831–7.
- Sendi P. Dealing with bad risk in cost-effectiveness analysis: the cost-effectiveness risk-aversion curve. *Pharmacoeconomics*. 2021;39:161–9.
- Sortino FA, Van Der Meer R. Downside risk. *J Portf Manag*. 1991;17:27–31.
- Sharpe WF. Mutual fund performance. *J Bus*. 1966;39:119–38.
- Sharpe WF. The sharpe ratio. *J Portf Manag*. 1994;21:49–58.
- Sendi P, Matter-Walstra K, Schwenkglenks M. Handling uncertainty in cost-effectiveness analysis: budget impact and risk aversion. *Healthcare*. 2021;9:1419.
- O'Brien BJ, Sculpher MJ. Building uncertainty into cost-effectiveness rankings: portfolio risk-return tradeoffs and implications for decision rules. *Med Care*. 2000;38:460–8.
- Graff Zivin J. Cost-effectiveness analysis with risk aversion. *Health Econ*. 2001;10:499–508.
- Graff Zivin J, Bridges JF. Addressing risk preferences in cost-effectiveness analyses. *Appl Health Econ Health Policy*. 2002;1:135–9.
- Elbasha EH. Risk aversion and uncertainty in cost-effectiveness analysis: the expected-utility, moment-generating function approach. *Health Econ*. 2005;14:457–70.

14. Abellan JM, Pinto JL, Mendez I, Badia X. Towards a better QALY model. *Health Econ.* 2006;15:665–76.
15. Wakker PP. Explaining the characteristics of the power (CRRA) utility family. *Health Econ.* 2008;17(12):1329–44.
16. Bridges J, Stewart M, King M, van Gool K. Adapting portfolio theory for the evaluation of multiple investments in health with multiplicative extension for treatment synergies. *Eur J Health Econ.* 2002;3:47–53.
17. Sendi P, Al MJ, Zimmermann H. A risk-adjusted approach to comparing the return on investment in health care programs. *Int J Health Care Finance Econ.* 2004;4:199–210.
18. Córdoba JC, Ripoll M. Risk aversion and the value of life. *Rev Econ Stud.* 2016;84:1472–509.
19. Lakdawalla DN, Phelps CE. Health technology assessment with risk aversion in health. *J Health Econ.* 2020;72:102346.
20. Plantinga A, de Groot JS. Risk-adjusted performance measures and implied risk attitudes. *J Perform Meas.* 2002;6:9–20.
21. Werner J. Risk aversion. In: Palgrave M (eds) *The new Palgrave dictionary of economics.* Palgrave Macmillan, London; 2008. pp. 1–6. https://doi.org/10.1057/978-1-349-95121-5_2741-1.
22. Bar-Shira Z, Just R, Zilberman D. Estimation of farmers' risk attitude: an econometric approach. *Agric Econ.* 1997;17:211–22.
23. Szpiro G. Measuring risk aversion: an alternative approach. *Rev Econ Stat.* 1986;68:156–9.
24. Cher D, Miyamoto J, Lenert L. Incorporating risk attitude into Markov-process decision models. *Med Decis Mak.* 1997;17:340–50.