ORIGINAL PAPER

A guide to extending and implementing generalized risk-adjusted cost-effectiveness (GRACE)

Darius N. Lakdawalla[1,2] · Charles E. Phelps[3,4]

Abstract

The generalized risk-adjusted cost-effectiveness (GRACE) model generalizes conventional cost-effectiveness analysis (CEA) by introducing diminishing returns to Health-Related Quality of Life (QoL). This changes CEA practice in three ways: (1) Willingness to pay (WTP) increases exponentially with untreated illness severity or pre-existing permanent disability, and WTP ends up lower for mild diseases but higher for severe diseases compared with conventional CEA; (2) Average treatment effectiveness should be adjusted for uncertainty in outcomes; and (3) The marginal rate of substitution between life expectancy and QoL varies with health state. Implementing GRACE requires new parameters describing risk preferences over QoL, the marginal rate of substitution between life expectancy (LE) and QoL, and the variance and skewness of treatment outcomes distributions. In this paper, we provide: (1) a generalized WTP threshold incorporating the possibility of permanent disability; (2) a simpler method to estimate the tradeoff rate between QoL and LE, eliminating the need to carry out treatment-by-treatment estimates; (3) a more-general method to adjust WTP for illness severity that permits non-constant relative risk-aversion in QoL; (4) a new approach to estimating risk-preferences over QoL, leveraging established empirical methods from "happiness" economics; and (5) a step-by-step guide for practitioners wishing to implement multi-period GRACE analyses.

Keywords Quality of life · Life expectancy · Health insurance · Public health insurance · Cost-effectiveness analysis

Introduction

Lakdawalla and Phelps (hereafter, LP) developed a generalized risk-adjusted cost-effectiveness (GRACE) framework [1, 2] that nests the traditional cost-effectiveness analysis (CEA) framework [3] as a special case. GRACE relaxes the

restriction of constant returns to health-related quality of life (QoL) imposed by traditional CEA.[1] This generalization produces three major implications for the proper conduct of CEA.

First, optimal willingness to pay (WTP) for health improvements increases exponentially as untreated disease severity rises. Cost-effectiveness thresholds become more generous for more severe illness, and lower for milder illnesses, perhaps differing by up to a factor of ten from lowest to highest severity [1]. For similar reasons, GRACE also shows that a given QoL gain is worth more to disabled than to otherwise-similar non-disabled persons.

Second, GRACE incorporates effects of uncertain health outcomes into health technology assessment (HTA). Reductions in outcome uncertainty provide value to risk-averse individuals

✉ Darius N. Lakdawalla
  dlakdawa@usc.edu

  Charles E. Phelps
  charles.phelps@rochester.edu

1   School of Pharmacy, Sol Price School of Public Policy, The Leonard D. Schaeffer Center for Health Policy and Economics, University of Southern California, Los Angeles, CA, USA

2   The National Bureau of Economic Research (NBER), Cambridge, MA, USA

3   University of Rochester, Rochester, NY, USA

4   Departments of Economics and Public Health Sciences, University Professor and Provost Emeritus, University of Rochester, Rochester, NY, USA

---

[1] Two recent studies assess whether there are diminishing returns to QoL. Attema et al. [4] provide an estimate of the elasticity of utility with respect to health of 0.8. Nielsen et al. [5] estimated the effects of disease severity on willingness to pay (WTP). Across their various specifications, the elasticity of WTP with respect to illness severity ranged from 0.12 to 0.30.

(and conversely). HTAs should include these effects, particularly in contexts with modest incremental average health gains.

Third, GRACE shows that people with lower QoL are more willing than those in better health to trade remaining life expectancy (LE) for more QoL. Conventional CEA, exemplified by the phrase "a QALY is a QALY…" assumes this tradeoff is independent of baseline QoL, sharply contrasting with population-based surveys [6–8].

In this analysis, we make five new contributions to the GRACE model, all of which help to clarify and broaden its real-world application. First, we generalize previous methods to show why how and why WTP can systematically increase as permanent disability worsens. Next, we simplify methods to estimate the tradeoff between LE and QoL, eliminating the requirement for the treatment-by-treatment estimation proposed by LP [1, 2]. Third, we present a novel method for estimating the risk-preference parameters required by GRACE. Fourth, we widen the range of risk preferences accommodated by the GRACE framework. Finally, we incorporate these findings into a step-by-step guide for practitioners seeking to conduct GRACE studies in both static and dynamic cost-effectiveness value assessments, extending the original static framework [1].

To develop these new ideas, we first summarize the original GRACE model [1, 2], and then present our first four contributions in Sect. "Simplification and expansion of GRACE". Section "Guide to using the GRACE method" provides a "handbook" for practitioners that clarifies specific implementation steps for using GRACE in a multiperiod setting.

Summary of the GRACE framework

LP consider an individual with expected period utility, $E[U(C)W(H)]$, where C is income available for consumption after medical spending is removed, and H is QoL, a stochastic index of health itself (not utility of health). Where all terms are shown in Box 1, the total value of a medical intervention (TVMI) from LP ([1], Eq. 16) is:

$$TVMI = K\omega_H R\{\mu_p\delta + \phi p_1\mu_B\epsilon\}. \tag{1}$$

The associated incremental generalized risk-adjusted cost-effectiveness ratio (IGRACER) decision rule is to adopt or reimburse the technology if ([1], Eq. 17):

$$\frac{\Delta C}{\{\mu_p\delta + \phi p_1\mu_B\epsilon\}} \leq K\omega_H R. \tag{2}$$

BOX 1. Definitions

$$C = \text{Income} - \text{Medical Spending} = \text{Consumption}$$

$$\Delta C = \textbf{Incremental cost of medical technology}$$

$$\omega_C = \frac{U'(C)C}{U(C)} = \text{elasticity of utility with respect to C}$$

$$\mathbf{K = \frac{C}{\omega_C} = \textbf{traditional cost-effectiveness threshold}}$$

$$H_{1S} = \text{QoL with untreated acute illness and (optionally) disability}$$

$$\mu_H \equiv E(H_{1S}), \text{average QoL with untreated acute illness and disability}$$

$$B = \text{stochastic gain in QoL from treatment in the sick state}$$

$$H_0 = \text{QoL in the baseline period, typically assumed to equal 1}$$

$$\mathbf{\omega_H = \frac{W'(H_0)H_0}{W(H_0)} = \textbf{elasticity of utility with respect to QoL (H) at } H_0}$$

$$R = \frac{W'(H_{1S})}{W'(H_0)}, \text{ disease-severity ratio}$$

H_{0d} = QoL in the baseline period inclusive of any permanent disability

H_{1W} = Health-related QoL in the well state, where $H_{1W} = H_{0d}$

$$d^* \equiv \frac{H_0 - H_{1W}}{H_0} = \text{relative quality of life loss from permanent disability}$$

$$\ell^* \equiv \frac{(H_0 - \mu_H)}{H_0}, \text{average \% QoL lost from untreated disease and from disability}$$

$$t^* \equiv \frac{(H_0 - E(H_{1S} + B))}{H_0}, \text{average \% QoL lost from treated disease and from disability}$$

μ_P = **Expected gain in probability of survival from treatment**

δ = **marginal rate of substitution (MRS) between life expectancy and QoL**

ϕ = **probability of acute illness in period 1**

p_1 = **probability of survival into period 1**

$\mu_B = E(B)$, **Expected gain in QoL from treatment**

ϵ = **Certainty-equivalent QALYs per average QALY gained**

$^\dagger r_H^*$ = Relative risk-aversion over QoL, evaluated at H_{1S}

$\dagger \pi_H^*$ = Relative prudence over QoL, evaluated at H_{1S}

$\Delta\sigma_{H_T}^2$ = (QoL Variance in Treatment Group) − (QoL Variance in Control Group)

$\Delta[\gamma_1 \sigma_H^{*3}]$ = (QoL Skewness in Treatment Group) − (QoL Skewness in Control Group)

Within Box 1, the parameters appearing directly in Eqs. [(1, 2)] are bolded, and the others are used to calculate the bolded terms.[2] Most of the Box 1 parameters are common to traditional CEA [3]. Exceptions unique to GRACE include the elasticity of QoL utility (ω_H), the disease-severity ratio (R), the certainty-equivalence ratio (ϵ), and the marginal rate of substitution between LE and QoL (δ). LP showed that all these new parameters—except δ—can be calculated using estimates of the variability of treatment outcomes (the $\Delta\sigma_H$ and $\Delta\gamma_1\sigma_H$ terms in Box 1) and of relative risk preferences over health (relative risk-aversion, r_H^* [9] and relative prudence, π_H^* [10]). Section "Simplification and expansion of GRACE"C provides empirical methods to estimate these risk-preference parameters.

GRACE nests traditional CEA as the special case of constant marginal utility from QoL. Constant marginal utility in QoL implies that $\omega_H = R = \epsilon = 1$ and $r_H^* = 0$.

---

[2] Within Box 1, the two terms marked with †, namely π_H^* and r_H^*, are evaluated at H_0 when used to estimate ψ and ρ, but evaluated at H_{1S} otherwise.

In contrast, GRACE allows for $0 < \omega_H \leq 1$, $R \geq 1$, $\epsilon \neq 1$, and $r_H^* \geq 0$. The disease-severity ratio, R, adjusts WTP to reflect the effects of diminishing returns to QoL—risk-averse people in worse QoL states place more value on a given QoL gain. The certainty–equivalence ratio, ϵ, adjusts the traditional quality-adjusted life-year (QALY) to account for QoL risk according to consumer risk preferences.

Simplification and expansion of GRACE

This section develops four improvements to the GRACE framework as originally presented in [1, 2], all of which either extend the range of situations where GRACE can be implemented or simplify its application.

A. Allowing disability in baseline health status

Begin with one of the most basic ideas in consumer theory—the marginal rate of substitution between two goods. From this notion, we develop a new way to determine WTP for medical technology.

Consider the cost-effectiveness threshold defined by traditional analysis. Define C_0 as period zero consumption and H_{1S} as (initially non-stochastic) QoL in the period one sick state. The marginal rate of substitution between C_0 and H_{1S} measures the period zero WTP for QoL improvement in the sick state. This can be thought of as the WTP for an insurance policy covering treatments that marginally improve QoL in the sick state. The traditional CEA framework presumes $V(C, H) = U(C)W(H)$, which yields the marginal rate of substitution (the ratio of marginal utilities of C and H):

$$\frac{dC_0}{dH_{1S}} = \frac{W'(H_{1S})U(C_0)}{U'(C_0)W(H_0)}. \tag{3}$$

Traditional CEA imposes the restriction that $W(H) = H$ and thus $W'(H) = 1$. It also presumes that $H_0 = 1$. With these restrictions, the traditional cost-effectiveness threshold is $K \equiv \frac{C_0}{\omega_C}$, where $\omega_C = \frac{U'(C_0)}{U(C_0)}C_0$ is the elasticity of utility with respect to consumption [11]. As discussed in LP ([1], p.3), K is the WTP for QoL improvement, and the QALY provides a vehicle for expressing all health improvements—including those resulting from LE gains—in terms of an equivalently valued QoL gain.

LP relax the restriction that $W(H) = H$ and allow for strictly concave $W(H)$. Also, they allow for stochastic health in the sick state: H_{1S} is random. To make the cost-effectiveness threshold non-stochastic, LP evaluate it at μ_H, the average QoL in the untreated period 1 sick state. This generalizes Eq. (3) as:

$$\frac{dC_0}{d\mu_H} = \left[\frac{C_0}{\omega_C H_0}\right]\omega_H\left[\frac{W'(\mu_H)}{W'(H_0)}\right]. \tag{4}$$

LP define $\omega_H \equiv \frac{W'(H_0)}{W(H_0)}H_0$, the elasticity of utility with respect to QoL, and $R \equiv \frac{W'(\mu_H)}{W'(H_0)}$, the "disease severity ratio." With these definitions, the generalized WTP threshold is simply $K_{GRACE} = K\omega_H R$. LP prove that this is the WTP for a generalized risk-adjusted QALY (GRA-QALY), given by $\mu_p\delta + \phi p_1\mu_B\epsilon$.

Here, we further generalize the LP result, allowing for permanent disability. Instead of requiring "excellent" period zero health, we assume instead that period zero health is $H_{0d} = H_0(1 - d^*)$, where $0 \leq d^* < 1$ is the percentage of QoL lost to permanent disability, as defined in Box 1.[3] The earlier LP model and traditional CEA study the special case where $d^* = 0$. In the more general case, Eq. (4) becomes:

$$K_{GRACE} = \frac{K\omega_H R}{H_0}\left[\frac{W(H_o)}{W(H_{0d})}\right]. \tag{5}$$

Appendix IA demonstrates that $\frac{W(H_{0d})}{W(H_0)} = (1 - d^*\psi)$, where the "disability ratio," ψ, is a new parameter that depends on ω_H, relative risk-aversion (r_H^*), and relative prudence (π_H^*) over QoL, all of which are already required by GRACE analyses. Compactly, the generalized WTP is:

$$K_{GRACE} = \frac{K\omega_H R}{H_0(1 - d^*\psi)}. \tag{6}$$

Appendix IA demonstrates that $\psi = 1$ when consumers are risk-neutral over QoL and $\psi < 1$ when they are strictly risk-averse. Therefore, weak risk-aversion implies $\psi \leq 1$, which ensures that K_{GRACE} is strictly positive and well-defined.

Appendix IC proves that people with permanent disabilities ($d^* > 0$) always exhibit greater WTP for QoL improvements than similar non-disabled people, and can sometimes even exhibit weakly greater WTP for LE improvements.[4] Intuitively, disability lowers the marginal utility of consumption, $V_C(C, H_{0d}) = U'(C)W(H_{0d})$ and thus increases the willingness to pay for health

---

[3] We assume throughout that $H_{1W} = H_{0d}$, so that permanent disability persists from period zero onwards. This generalization also requires that H_{1S} include permanent disability; therefore, analysts must calculate H_{1s} to reflect the effects of untreated acute illness and any pre-existing permanent disability.

[4] As discussed in Appendix IC, these claims rely on constant relative risk-aversion over QoL. If risk preferences instead vary with QoL, a variety of complex effects may arise, and the relationship must be studied empirically.

improvements. This leads to unambiguously higher WTP for QoL among those with disabilities. Some ambiguity arises for LE, because the lower marginal utility of consumption competes with a decline in period utility, $U(C)W(H_{1S})$, also caused by disability.

With permanent disability, Eq. (1) for the total value of a medical intervention (TVMI) generalizes to:

$$TVMI' = \frac{K\omega_H R}{H_0(1 - d^*\psi)}\{\mu_p\delta + \phi p_1\mu_B\epsilon\}. \tag{7}$$

The total marginal value of life-extension, μ_p, is $\left(\frac{K\omega_H R}{H_0(1-d^*\psi)}\right)\delta$, which is the WTP for QoL gains $\left(\frac{K\omega_H R}{H_0(1-d^*\psi)}\right)$ multiplied by the marginal rate of substitution between LE and QoL. Permanent disability unambiguously increases $\frac{1}{1-d^*\psi}$ and thus pushes up WTP for QoL gains. On the other hand, disability lowers the marginal rate of substitution between LE and QoL, δ, because an additional period of life becomes less valuable when period utility falls.[5]

Appendix IC characterizes the net effect: defining B as the stochastic QoL benefit of medical technology, permanent disability weakly increases WTP for LE if and only if $\frac{E(W(H_{1S}+B))}{W(H_{0d})}$ rises with disability. In other words, if a given disability lowers utility proportionally more in healthy states than sick states, it will increase the WTP for LE; heuristically, this requires that acute illness "crowds out" some of disability's QoL effects. For example, this will be true if utility exhibits CRRA and disability lowers QoL by a constant percentage in both healthy and sick states.

To characterize the condition more generally, define t^* as the percentage QoL loss after treatment, inclusive of disability, so that $E(H_{1S} + B) = H_0(1 - t^*)$, as in Box 1. Appendix IC proves that—under CRRA utility—disability weakly increases WTP for LE if and only if $\frac{\partial t^*}{\partial d^*} \leq \frac{1-t^*}{1-d^*}$. This is always satisfied for curative therapies where $t^* = d^*$, because disability reduces QoL by the same percentage in both states. Imperfect therapies make this restriction non-trivial.

This analysis explains why standard methods of CEA discriminate against disabled persons, since the standard approach assumes $d^* = 0$. Holding $d^* = 0$ implies that a QALY is *always* worth less for disabled persons than for otherwise-similar non-disabled persons. This problem has led to several ad hoc attempts to resolve this inconsistency [12, 13]. Such ad hoc adjustments are unfounded and perhaps even unnecessary. Under GRACE, disability always increases the WTP for QoL gains and can weakly increase

the WTP for LE gains if $\frac{\partial t^*}{\partial d^*} \leq \frac{1-t^*}{1-d^*}$. Analysts can use this condition to determine the frequency with which disability leaves WTP for LE at least unchanged. The U.S. Affordable Care Act forbids use of CEA methods that discount the value of life for disabled people, both for assessing treatment value and determining Medicare coverage.[6] GRACE demonstrates how WTP for life-extension can rise with disability and may help provide a framework for conducting CEA compatible with current law.

B. Simplified estimation of the marginal rate of substitution (MRS)

We now turn to a second improvement upon LP, showing how to recover the marginal rate of substitution, δ, from other GRACE parameters. This obviates the need for disease-specific estimation of δ. LP formally defined this MRS as:

$$\delta \equiv \frac{\phi[E[W(H_{1S} + B)] + (1 - \phi)W(H_{1W})]}{W'(\mu_H)} \equiv \frac{V}{W'(\mu_H)}. \tag{8}$$

Recall H_{1S} is the stochastic period 1 level of untreated health in the "sick" state, B is the stochastic QoL benefit arising from medical interventions, and H_{1W} represents non-stochastic health when no acute illness occurs. The numerator, V, is the gain in expected utility produced by the marginal increase in survival probability, and the denominator, $W'(\mu_H)$, describes the marginal gain in utility arising from improved QoL in the untreated sick state. Since disability influences μ_H, H_{1S}, and H_{1W}, δ varies with disability. For the same reason, δ varies with the severity of untreated illness (μ_H). This is why LP specified that δ must be estimated on a disease-by-disease basis.

Appendix IB proves that TVMI in Eq. (1) can also be written as[7]:

$$TVMI = K[\mu_p\rho H_0 + \phi p_1\omega_H R\mu_B\epsilon]. \tag{9}$$

The new parameter, $\rho \equiv \frac{\phi[E[W(H_{1S}+B)]+(1-\phi)W(H_{1W})]}{W(H_0)} = \frac{V}{W(H_0)}$, is the ratio of period 1 expected QoL utility to baseline period QoL utility.

---

[5] See the discussion around Eq. (8) below.

Table 1 Values of disease-severity multiplier (R) by disease severity (ℓ^*) and relative risk-aversion (r_H^*)

ℓ^*	r_H^*				
	0	0.25	0.5	0.75	1
0	1	1	1	1	1
0.1	1	1.03	1.05	1.08	1.11
0.3	1	1.09	1.2	1.31	1.43
0.5	1	1.19	1.41	1.62	2
0.7	1	1.35	1.83	2.47	3.33
0.9	1	1.78	3.15	5.61	10

In general, $0 < \rho \leq 1$, where equality holds only for perfect cures coupled with $d^* = 0$. Furthermore, $H_{1W} = H_0(1 - d^*)$ and $E(H_{1S}) = H_0(1 - \ell^*)$, where ℓ^* is the average percentage health loss from the acute illness and any pre-existing permanent disability (see Box 1).[8] Both H_{1W} and H_{1S} are measured on a [0, 1] scale for an appropriate index of health-related QoL.

We now show that ρ can be estimated simply from standard measures of treatment effectiveness, such as from randomized clinical trials (RCTs) or similar evaluation methods, combined with measures of risk-preferences that are not disease-specific. This implies that δ can be estimated without disease-specific preference studies.

Appendix IB demonstrates that $\delta = \frac{\rho H_0}{\omega_H R}$, using the definitions of ρ, δ, and ω_H. Moreover, ρ can be expressed using Taylor series expansions around H_0, the baseline health level. Appendix IB describes ρ as the ϕ-weighted sum of two parts, one for the "sick" (acute illness) state (S) and the other for the "well" state (W), both of which should incorporate effects of any permanent disabilities:

$$\rho = \phi\rho_S + (1 - \phi)\rho_W \qquad (10)$$

Now define t^* such that $E(H_{1S} + B) = H_0(1 - t^*)$, so t^* is the average percentage reduction in QoL *after* treatment. For example, if *ex ante* health is perfect ($H_0 = 1$), and average post-treatment QoL is 0.75, then $t^* = 0.25$. If the treatment is perfect, then $t^* = d^*$, and if $d^* = 0$, then $t^* = 0$ also.

Appendix IB shows that:

$$\rho_S = 1 - \omega_H\left[t^* + \frac{1}{2}r_H^*t^{*2} + \frac{1}{6}r_H^*\pi_H^*t^{*3} + \ldots\right]. \qquad (11)$$

In Eq. (11), health losses are "discounted" by ω_H, which reflects rates of diminishing returns to health. Similarly, the full Taylor Series expansion for the "well state" portion of ρ is:

$$\rho_W = 1 - \omega_H\left[d^* + \frac{1}{2}r_H^*d^{*2} + \frac{1}{6}\pi_H^*r_H^*d^{*3} + \ldots\right] \qquad (12)$$

Thus, $\rho_W = 1$ when $d^* = 0$. Equation (12) completes the characterization of Eq. (10). Once the utility parameters ω_H, r_H^* and π_H^* are estimated, t^* and d^* can be estimated from standard QoL outcomes studies. With these new results, analysts can estimate ρ without the need of disease- or treatment-specific studies that would otherwise be required when using the definition of δ in Eq. (8).

C. Estimating willingness to pay for non-constant relative risk-aversion

LP demonstrate that WTP for generalized risk-adjusted QALYs is $K_{GRACE} = K\omega_H R$. Prior research demonstrates that $K = \frac{C}{\omega_C}$ [11]. The general economics literature suggests that $0.3 \leq \omega_C \leq 0.5$ [11, 14, 15]. This implies (approximately) that for consumption level C, $2C \leq K \leq 3C$.

In general, $0 < \omega_H \leq 1$ and $R \geq 1$, where R grows with illness severity. LP showed how to recover estimates of R under the assumption of constant relative risk-aversion (CRRA) over QoL. We next explain why and how CRRA may incorrectly restrict possible values of R, and then present a more-general method for estimating R under non-CRRA utility.

1. WTP under constant relative risk-aversion

Table 1 gives a representative set of values for R under CRRA utility and different combinations of r_H^* and untreated illness severity (ℓ^*).

Table 1 illustrates how willingness-to-pay for QoL improvements rise under CRRA as both relative risk aversion (r_H^*) and severity of illness (ℓ^*) increase.[9]

To pursue Table 1's implications, note that traditional CEA presumes that $\omega_H = 1$, and $r_H^* = 0$. For illustrative purposes, assume that $\omega_C = 0.33$, a value within the range of prior empirical literature [11, 14, 15]. In this case, $K = 3C$ at all levels of disease severity. For example, at annual consumption of $C = \$50,000$, the WTP per QALY under traditional CEA would be $K = \$150,000$ for all disease levels.

By comparison, $K_{GRACE} = \frac{K\omega_H R}{H_0(1 - d^*\psi)}$. For simplicity, let $d^* = 0$ and $H_0 = 1$. Now, suppose that $\omega_H = 0.5$ instead of 1.0, and ω_C remains as 0.33. Thus, $K\omega_H = 1.5C = \$75,000$. To calibrate R, note that $\omega_H = 1 - r_H^*$ under CRRA, and thus $\omega_H = 0.5$ implies

---

[8] In [1, p. 9], due to a typographical error, this was defined as $\ell^* \equiv H_0 - \mu_H$. Box 1 correctly defines $^* \equiv \frac{H_0 - \mu_H}{H_0}$, a distinction that matters only in exceptional cases where analysts wish to set $H_0 \neq 1$.

[9] For comparison, Nielson et al. [5] estimated that the elasticity of R with respect to illness severity was (at their midpoint) 0.2. This is approximately consistent with the values in Table 1 when $r_H^* = 0.25$.

$r_H^* = 0.5.$[10] From Table 1, $r_H^* = 0.5$ implies values for R ranging from 1.0 (for $\ell^* \approx 0$) to 3.15 (for $\ell^* = 0.9$). This means that $\$75,000 \leq K_{GRACE} \leq \$236,250.$ Similarly, when $\omega_H = 0.75$ (hence $r^* = 0.25$ with CRRA), $\$112,500 \leq K_{GRACE} \leq \$200,270.$ When $\omega_H = 0.25$ (hence $r_H^* = 0.75$), $\$37,500 \leq K_{GRACE} \leq \$210,377.$ If included, permanent disability ($d^* > 0$) would increase WTP relative to these estimates.

As these examples demonstrate, the severity gradient of R rises as r_H^* rises. At the same time, however, the tight linkage between ω_H and r_H^* imposed by CRRA restricts the space of possible values for r_H^*. The implications of non-CRRA utility range from minimal to consequential, depending on specific parameter estimates. We pursue this issue next.

2. Relaxing the CRRA assumption

The Hyperbolic Absolute Risk Aversion (HARA) family of utility permits all forms of risk-aversion, including not only CRRA, but also increasing and decreasing relative risk aversion (IRRA and DRRA).[11] We use a slightly simplified HARA utility function (see Appendix II):

$$U(H) = \left[\frac{1-\gamma}{\gamma}\right] Z^\gamma, \tag{13}$$

where $Z = H + \eta$. The parameter η has the same units of measurement as H, and can be positive or negative. We restrict the domains of H and η such that $Z \geq 0$, and we specify that $U(H) = 0$ for $Z = 0$. For this simplified HARA model, Appendix II proves that:

$$r_H^* = (1-\gamma)\left(\frac{H}{Z}\right) = (1-\gamma)\left[\frac{H}{H+\eta}\right]. \tag{14}$$

When $\eta = 0$, HARA simplifies to CRRA utility, because $H = Z$ and thus $r_H^* = (1-\gamma)$, which is constant. However, when $\eta > 0$, $r_H^* < (1-\gamma)$, and when $\eta < 0$, $r_H^* > (1-\gamma)$, and can exceed 1.0 for sufficiently small values of γ and sufficiently large absolute values of η.[12] This expands the realm of possible values of r_H^* beyond those available with the CRRA restriction.

Similarly (see Appendix II), the elasticity of utility with respect to H is:

$$\omega_H = \gamma\left[\frac{H}{H+\eta}\right] = \gamma\left(\frac{H}{Z}\right). \tag{15}$$

Taking the ratio of (14) to (15) yields an important insight: Since $\left(\frac{H}{Z}\right)$ cancels out:

$$\frac{r_H^*}{\omega_H} = \frac{1-\gamma}{\gamma}. \tag{16}$$

Therefore,

$$\gamma = \frac{\omega_H}{r_H^* + \omega_H}. \tag{17}$$

Hence, $0 < \gamma < 1$ since both r_H^* and ω_H are positive for risk-averse individuals.

Identifying γ also provides a basis for estimating *all* higher-order risk parameters in HARA utility once r_H^* and ω_H (and hence γ) have been estimated. Phelps and Cinatl [15] prove that for HARA utility:

$$\pi^* = \frac{2-\gamma}{1-\gamma}r^*. \tag{18}$$

and:

$$\tau^* = \left[\frac{3-\gamma}{1-\gamma}\right]r^*. \tag{19}$$

The general form, where ζ_j^* is the jth relative risk preference parameter, is:

$$\zeta_j^* = \frac{j-\gamma}{1-\gamma}r^*. \tag{20}$$

These higher-order risk parameters enter the Taylor-Series expansions for the severity-of-illness ratio R, the disability ratio ψ, and the estimation of ρ outlined above, so having a convenient way to estimate them is useful. This generalizes previous methods that relied upon CRRA to estimate R from ℓ^*.[13] With these derivations for HARA utility, employing GRACE requires only two QoL parameters—ω_H and r_H^*—the estimation of which we turn to in the next section.

Appendix III illustrates that higher-order relative risk preference parameters may be underestimated by 50% or more for some HARA utility functions that fail to satisfy CRRA. Appendix III also shows that: (1) relative bias is exacerbated if the degree of relative risk-aversion is lower; and (2) bias becomes more consequential if QoL outcomes possess highly non-normal distributions, in which

---

[10] See Eqs. (14) and (15) and surrounding discussion for explanation of why $\omega_H = 1 - r_H^*$ under CRRA.

[11] For an expanded discussion of HARA utility, see [15].

[12] When $\eta < 0$, this formulation closely resembles the widely-used Stone-Geary utility function, where the η parameter represents minimum values of consumption before any utility is produced.

[13] In CRRA utility, the higher order risk preferences are determined by $\pi^* = r^* + 1, \tau^* = r^* + 2, \ldots etc.$

case relative prudence and other higher-order risk parameters become more important.[14] These results further emphasize the importance of understanding whether utility is CRRA or not, and if not, properly estimating the relative risk preference parameters.

D. Parameter estimation using "happiness" models

LP [1] proposed estimating relative risk preference parameters using discrete choice experiments (DCEs) of the sort that prior research [16] has employed for recovering relative risk preferences over consumption. This method remains viable. Once analysts have estimated r_H^* and π_H^* using standard DCE methods, they can exploit the properties of HARA utility to recover γ [see Appendix Eq. (72)], and they can use Eq. (16) to recover ω_H. Best-practices for the design and conduct of discrete choice experiments have been presented elsewhere [17].

However, discrete choice approaches tend to rely on relatively complex survey instruments. In some cases, analysts may wish to field shorter and simpler survey instruments to reach larger samples of respondents and achieve higher completion rates. For such contexts, we offer an additional strategy that exploits the empirical "happiness" literature pioneered by Easterlin [18]. Specifically, we show how to use the empirical happiness approach to estimate ω_H and r_H^*, under the assumption of HARA utility. Analysts can then exploit results of Sect. "Simplification and expansion of GRACE"C to recover the full set of risk preference parameters from ω_H and r_H^*, estimating γ using Eq. (17) and higher-order terms using Eqs. (18, 19, 20).

Following Easterlin [18], we propose that analysts estimate "utility" by using respondents' self-reported happiness on an ordinal scale ("*Happy*"), and the use of higher-order regression terms to capture the curvature of preferences more fully. Furthermore, we propose the collection of data on respondents' annual consumption (C) and some suitable index of health-related QoL (H). Now, exploiting the "translog utility" estimation approach of Christensen, Jorgenson and Lau [19], consider the following regression model relating self-reported happiness to self-reported levels of health (H_i) and consumption (C_i) for individual i:

$$
\begin{aligned}
\ln\left(Happy_i\right) &= \beta_1 \ln\left(H_i\right) + \frac{1}{2}\beta_2 \left(\ln\left(H_i\right)\right)^2 \\
&\quad + \beta_3 \ln\left(C_i\right) + \frac{1}{2}\beta_4 \left(\ln\left(C_i\right)\right)^2 + \epsilon_i.
\end{aligned}
\tag{21}
$$

Suppressing individual-specific subscripts, i, Eq. (21) provides a second-order approximation of any sufficiently differentiable utility function [19]. The above formulation suppresses interaction terms between $\ln(C)$ and $\ln(H)$, but these could be added if desired. In what follows, we develop our results in terms of H, but exactly analogous developments can be made to recover relative risk-aversion parameters over consumption, C.

Noting that $\frac{\partial \ln(Happy)}{\partial \ln(H)} = \omega_H$, from Eq. (21):

$$
\omega_H(H) = \beta_1 + \beta_2 \ln(H).
\tag{22}
$$

This implies the estimator, $\widehat{\omega_H}(H) = \widehat{\beta}_1 + \widehat{\beta}_2 \ln(H)$, so ω_H varies with H. Equation (16) then implies that r_H^* will vary with H. And, Eqs. (18, 19, 20) then imply that π_H^* and the other higher-order risk-preference parameters will vary with H also. Therefore, this estimation framework allows risk preferences to vary with the level of health (H), instead of requiring CRRA.

We now turn to estimation of r_H^* using the "happiness" equation approach, details of which appear in Appendix IV. To begin, we apply a general rule for any sufficiently differentiable utility function. Define $\epsilon^\omega \equiv \frac{\partial \omega_H}{\partial H} \frac{H}{\omega_H}$ as the elasticity of ω_H with respect to H. Suppressing the various dependencies on H, Appendix IV proves that:

$$
\epsilon^\omega = 1 - \left(r_H^* + \omega_H\right).
\tag{23}
$$

So,

$$
r_H^* = \left(1 - \omega_H\right) - \epsilon^\omega.
\tag{24}
$$

When utility is CRRA, $\epsilon^\omega = 0$. Thus, Eq. (24) implies that $\omega_H = 1 - r_H^*$ and $0 < r_H^* < 1$. When utility has declining relative risk aversion (DRRA), Eq. (16) implies that $\epsilon^\omega < 0$. Thus, Eq. (23) implies that $r_H^* + \omega_H > 1$. The converse is true if utility exhibits increasing relative risk aversion (IRRA).

With an estimate of ω_H in hand, we can use Eq. (24) to recover r_H^*, provided we can estimate ϵ^ω. Note that since $\omega_H = \beta_1 + \beta_2 \ln(H)$, then $\frac{\partial \omega_H}{\partial H} = \frac{\beta_2}{H}$. From this,

$$
\epsilon^\omega(H) = \frac{\partial \omega_H}{\partial H}\left(\frac{H}{\omega_H}\right) = \frac{\beta_2}{\omega_H} = \frac{\beta_2}{\beta_1 + \beta_2 \ln(H)}.
\tag{25}
$$

The empirical analog, $\widehat{\epsilon^\omega}(H) = \frac{\widehat{\beta}_2}{\widehat{\beta}_1 + \widehat{\beta}_2 \ln(H)}$, varies with H unless $\widehat{\beta}_2 = 0$. With estimates of ω_H and ϵ^ω in hand, we can now recover r_H^* from Eq. (24) as:

$$
\widehat{r}^*(H) = 1 - \widehat{\omega}_H(H) - \widehat{\epsilon^\omega}(H) = 1 - \left(\widehat{\beta}_1 + \widehat{\beta}_2 \ln(H)\right) - \frac{\widehat{\beta}_2}{\widehat{\beta}_1 + \widehat{\beta}_2 \ln(H)}.
\tag{26}
$$

If $\beta_2 < 0$ (DRRA utility) then r_H^* can exceed $\left(1 - \widehat{\omega}_H\right)$. For relatively small values of $\widehat{\beta}_1$ and sufficiently large negative values of $\widehat{\beta}_2$, $\widehat{r}_H^*$ can exceed 1.0. This widens the potential range for values of $\widehat{r}^*(H)$ compared with the stricter requirements of CRRA utility.

---

[14] For normal distributions, the risk terms above variance all equal zero.

Using estimates of ω_H^* and r_H^*, we can return to the HARA utility structure (Section "Simplification and expansion of GRACE"C2) to estimate γ [using Eq. (17)] and all higher-order relative risk preference parameters [using Eqs. (18, 19, 20)]. In turn, analysts can estimate R using these parameters and the formula for R from ([1], Eq. 19), reproduced here for convenience:

$$R = \left\{ 1 + r_H^* \ell^* + \frac{1}{2} r_H^* \pi_H^* \ell^{*2} + \frac{1}{6} r_H^* \pi_H^* \tau_H^* \ell^{*3} + \dots \right\}. \quad (27)$$

This method enables more robust estimates of R, compared with Table 1, which relies on the restrictive CRRA assumption.

Guide to using the GRACE method

GRACE calls for practitioners to introduce new information into HTAs. This section presents a guidebook for carrying out these tasks. It includes two sections—things that must be done only once, and things that must be done for every new HTA.

A. Things done only once

In the language of biopharmaceutical companies, these are "above-brand" tasks. They estimate parameters describing people's attitudes towards QoL risk. Practitioners who analyze specific healthcare interventions need not perform these tasks for each new HTA.

1. One-time parameter estimates

These "one-time" parameters represent a full complement of measures characterizing consumer attitudes towards QoL risk. They include:

1. The elasticity of utility with respect to health (ω_H). Conceptually, $0 < \omega_H < 1$ if marginal utility from improved health is positive but declining.
2. Relative risk aversion (r_H^*), related to variance of treatment outcomes. This parameter determines how the marginal utility of QoL changes as QoL itself changes.
3. Relative prudence (π_H^*), related to skewness of treatment outcomes. This parameter determines how r_H^* changes as H changes.
4. Optionally, relative temperance (τ_H^*), related to excess kurtosis of treatment outcomes. This parameter determines how π^* changes as H changes.

Technically, LP rely on relative risk-aversion parameters evaluated at μ_H, the average QoL in the untreated sick state, which will vary across disease

states.[15] Under CRRA, this subtlety can be ignored. However, analysts wishing to relax the CRRA assumption would ideally estimate relative risk preferences for varying degrees of QoL to create a library of such estimates applicable to a range of diseases.

Section "Simplification and expansion of GRACE"D (along with Appendices II and IV) presents two methods for recovering relative risk preference parameters: (1) estimating r_H^* and π_H^* via discrete choice experiments and recovering the remaining parameters; or (2) estimating ω_H and r_H^* from "happiness economics" regression parameters and features of HARA utility functions.

2. Estimating severity-adjusted WTP

Estimating risk- and severity-adjusted WTP requires the preference parameters described above and estimates of disease severity. Recall from Sect. "Simplification and expansion of GRACE"A that in the conventional model, where C is annual consumption and ω_C is the elasticity of utility with respect to C [11]:

$$K = \frac{C}{\omega_C}. \quad (28)$$

In the GRACE framework:

$$K_{GRACE} = \frac{K \omega_H R}{H_0 (1 - \psi d^*)}. \quad (29)$$

In addition to the measure of health loss from disability, d^*, this adds three new risk parameters—ω_H, R and the new parameter ψ, which can be estimated using risk parameters already required (see Appendix IA). As shown in Eq. (10), the parameter R depends on relative risk preferences, and the severity of untreated disease (ℓ^*). "Happiness economics" methods (Section "Simplification and expansion of GRACE"D) permit estimation of the key parameters ω_H and ω_C from the same data set, if desired.

Severity (untreated QoL) is already estimated for many diseases of interest [20]. We presume that, over time, tables reporting disease severity for various diseases (ℓ^*) will be created by public payers or health technology assessment bodies, such as NICE in the UK or CMS in the US. For example, just as CMS has calculated hospital payment levels using Diagnosis-Related Groups, it may wish to measure disease-severity within a similar taxonomy. Those disease-severity values, com-

---

[15] As noted in Appendix IA and B, the estimation of δ and ψ require relative risk preference parameters evaluated at H_0.

bined with population estimates of risk-preferences, allow estimation of R, ω_H, and ultimately the "system-wide" threshold of $K_{GRACE} = \frac{K\omega_H R}{H_0(1-\psi d^*)}$, for various levels of disease and disability severity. These are "done once" and need not be undertaken by practitioners evaluating specific medical interventions.

In standard CEA, medical technologies are considered "welfare-improving" if incremental costs ($\Delta Cost$) relative to incremental benefits ($\Delta QALYs$) are less than society's WTP, normally expressed as:

$$K \equiv \frac{C}{\omega_C} \geq \frac{\Delta Cost}{\Delta QALYs}. \tag{30}$$

GRACE modifies this so that:

$$K_{GRACE} \equiv \frac{K\omega_H R}{H_0(1-\psi d^*)} \geq \frac{\Delta Cost}{\Delta(GRA\text{-}QALYs)}, \tag{31}$$

where the generalized risk-adjusted QALY (GRA-QALY) is $\{\mu_p \delta + \phi p_1 \mu_B \epsilon\}$.

B. Parameters that must be estimated for each new treatment

Both R and GRA-QALYs themselves contain parameters that vary by treatment intervention, listed next:

1. *The average incremental gain in QALYs compared with the current "best practice treatment.* GRACE calls this μ_B ("B" for "Benefit"). Current studies routinely estimate μ_B, the difference between average QoL outcomes for the new treatment (T) and its comparison (C).

2. *The variance of patient outcomes, both for "T" and "C."* GRACE uses *the difference* between these variances, described in the model as $\Delta\sigma^2 \equiv \sigma_{H_T}^2 - \sigma_{H_C}^2$, where $\sigma_{H_T}^2$ and $\sigma_{H_C}^2$ are the QoL variance in the treatment and comparator groups. Reductions in variances of treatment options provide additional value beyond mean health gains, and conversely.

3. *The skewness of patient outcomes, both for "T" and "C".* This requires estimating Pearson's skewness parameters, γ_{1H_T} and γ_{1H_C}, for both treatment and comparator populations.[16] Where (generally) $\sigma^3 = [\sigma^2]^{\frac{3}{2}}$, GRACE uses $\Delta Skewness = \Delta[\gamma\sigma^3] \equiv \gamma_{1H_T}\sigma_{H_T}^3 - \gamma_{1H_C}\sigma_{H_C}^3$, where the subscripts H_T and H_C represent (respectively) the treatment and comparator populations. Increases in positive skewness add value for any given change in variances of outcomes (and conversely).

4. *Optionally, the kurtosis of outcomes in treatment and comparison groups.* Kurtosis magnifies effects of variance, so reductions in excess kurtosis have independent value in addition to other components of value in

GRACE. GRACE uses the difference in kurtosis values, parallel to the measures of skewness. Estimating kurtosis requires large sample populations, potentially infeasible in some clinical studies.[17]

5. *Estimates of the relative QoL loss from disease, both before and after treatment.* GRACE measures these on a percentage basis, relative to *ex ante* health states that obtain before the illness occurred.

   a. *Loss in untreated states*: This parameter ℓ^* (defined in Box 1) represents percentage losses in QoL suffered in the untreated states. This includes the QoL effects of acute illness and of any permanent disabilities. Diseases with no QoL consequences would imply $\ell^* = 0$; one that lowers QoL by 50% from its baseline state would involve $\ell^* = 0.5$, and so on. Many estimates of ℓ^* are already available from CEA studies [21]. This parameter is disease-specific, but not treatment-specific.

   b. *Loss in treated states:* This parameter t^* (defined in Box 1) reflects percentage losses in QoL remaining after diseases are treated. As with ℓ^*, this includes the QoL effects of treated acute illness and of any permanent disabilities. If treatment always completely cures patients, $t^* = 0$.[18] If post-treatment states feature QoL that lies 30% below the *ex ante* (pre-disease) QoL level (H_0), then $t^* = 0.3$, etc. These values must be estimated for each disease/treatment pair. As explained in Sect. "Simplification and expansion of GRACE"B, they are used to estimate the MRS between QoL and LE in GRACE.

   c. *Permanent disability loss:* This parameter d^* (defined in Box 1) reflects the percentage loss in QoL created by permanent disabilities that predated the arrival of the period one acute illness. This is an optional term that analysts may wish to consider when studying health interventions among people with disabilities. Under GRACE, people with disabilities always exhibit higher WTP for QoL improvements and sometimes exhibit equal or higher WTP for LE gains (see Sect. "Simplification and expansion of GRACE"A).

   d. *Ex ante QoL level:* H_0 measures QoL prior to onset of relevant disease in populations of interest. Typically, analysts will presume $H_0 = 1$.

6. *The average increase in survival probability to period 1, compared with the control therapy, described in GRACE*

---

[16] The variance of the estimated Pearson Skewness parameter in normal populations is approximately 6/n for sufficiently large samples.

[17] The variance of estimated excess kurtosis in normal samples is approximately 24/n for sufficiently large samples.

[18] This is true when $d^* = 0$. When $d^* > 0$, $t^* = d^*$.

as μ_p. The approach to calculating long-term survival gains differs depending on the stationarity of QoL distributions; we provide guidance on this issue in Sect. "Guide to using the GRACE method"C below.

7. *The probability of surviving to period 1, described in GRACE as p_1.* This is the level of survival probability, rather than its change.

8. *The population incidence of the disease of interest, ϕ.* This parameter is typically estimated in epidemiological burden of illness studies.

Traditional cost-effectiveness and comparative-effectiveness studies estimate Parameter 1, the mean gain in QoL due to treatment. Nearly all such studies can be used to recover Parameters 2, 3, and 4, simply by recording variance, skewness, and kurtosis of QoL outcomes for patients in treatment and control groups. Such studies likely already produce estimates for parameters listed in 5, which concern QoL levels in treated, untreated, and pre-onset states. The same is true for survival parameters 6 and 7, also routinely estimated in current cost-effectiveness studies. Finally, existing burden of illness studies likely already estimate disease incidence, parameter 8.

The risk-preference parameters from Sect. "Guide to using the GRACE method"A.1 and estimates of ℓ^* (item 5a in the list above) together allow estimates of R using Eq. (10). Next, the list of intervention-specific parameters above, combined with risk-preference parameters in Sect. "Guide to using the GRACE method"A, imply intervention-specific values of ρ, which can be calculated using Eqs. (23, 24, 25) above.

Next, parameters in Sect. "Guide to using the GRACE method"A and the list above imply intervention-specific values of ϵ, which can be calculated from ([1], Eq. 9), reproduced here:

$$\epsilon \approx 1 + \left[\frac{1}{\mu_B}\right]\left[-\frac{1}{2}r_H^*\left(\frac{1}{\mu_H}\right)\Delta\sigma_H^2 + \frac{1}{6}\pi_H^* r_H^*\left(\frac{1}{\mu_H}\right)^2\Delta\left[\gamma_1\sigma_H^3\right] + \dots\right].$$

(32)

When there are no differences in variance and skewness of QoL in the treatment and comparator groups, then $\epsilon = 1$, and no need exists to adjust mean treatment benefits. Improvements (worsening) of variance add to (subtract from) overall benefit, and increases (decreases) in skewness add to (subtract from) mean benefits.

C. Estimating net monetary benefit and cost-effectiveness

The original GRACE formulation employed a static two-period setting, in which patients are healthy in period zero and potentially sick in period one. Nonetheless, this static framework can handle dynamic treatment contexts and models (e.g., Markov models). We demonstrate this for two different cases.

Begin with the more-general case in which the means, variances, and/or skewness in QoL outcomes vary over time. Consider an assessment of value over periods $n = 1, 2, \dots, N$. The following parameters would then need to be estimated on a period-by-period basis: μ_{Bn} (mean QoL benefit in period n), $\Delta\sigma_{Hn}^2$ (difference in the variance of QoL benefit in period n), $\Delta\left[\gamma\sigma_H^3\right]_n$ (difference in skewness of QoL benefit in period n), ℓ_n^* (percentage QoL loss from untreated disease in period n, and t_n^* (percentage QoL loss from treated disease in period n).[19] Moreover, since ρ, δ, and ϵ also depend on treatment outcomes, they would be indexed as ρ_n, δ_n, and ϵ_n. Since R depends on ℓ^*, which might vary over time, it too is indexed as R_n. Define ΔC_n as incremental treatment cost incurred in period n, which is equivalent to the incremental reduction in non-health period n consumption, C_n. Finally, define μ_{pn} as the average increase in the probability of surviving from period $n-1$ to period n, define p_n as the post-treatment probability of surviving from period $n-1$ to period n, define Π_n as the cumulative post-treatment probability of surviving from period zero to period n.[20] Defining β as the one-period discount factor and recalling $\delta_n \equiv \frac{\rho_n H_0}{\omega_H R}$, the net monetary benefit in the dynamic setting is:

$$\text{NMB}_0 = \sum_{n=1}^{N}\beta^n\Pi_{n-1}\left\{\left[\frac{K\omega_H R_n}{H_0(1-d^*\psi)}\right]\left[\mu_{pn}\delta_n + \phi p_n\mu_{Bn}\epsilon_n\right] - p_n\Delta C_n\right\} - \Delta C_0.$$

(33)

This expresses net monetary benefit discounted in terms of period zero consumption, NMB_0.

A suitable incremental generalized risk-adjusted cost-effectiveness ratio (IGRACER) decision rule can be calculated as:

$$\frac{\Delta C_0 + \sum_{n=1}^{N}\beta^n\Pi_{n-1}\left\{p_n\Delta C_n\right\}}{\sum_{n=1}^{N}\left\{\beta^n\Pi_{n-1}\left(\frac{R_n}{R_1}\right)\left[\mu_{pn}\delta_n + \phi p_n\mu_{Bn}\epsilon_n\right]\right\}} \leq \frac{KR_1\omega_H}{H_0(1-\psi d^*)}.$$

(34)

Note that analysts can freely choose any period $j = 1, \dots, N$ to normalize the WTP threshold on the right-hand side, resulting in this more general expression:

---

[19] While we suppress the extra notation, note that variance and skewness in both the treatment and comparator arms should also be estimated in each period, in addition to the differences in variance and skewness across arms. In addition, we presume that permanent disability is fixed over time and that any variation in it could be represented using additional time-series variation in ℓ_n^* and t_n^*.

[20] This formulation presumes that QoL improvements follow or coincide with life expectancy gains. In cases where QoL improvements precede life expectancy gains, p_n and Π_n should instead measure pre-treatment life expectancy.

Table 2 Summary of GRACE parameters to be estimated

Parameter		Fixed (F), or varying over time (T)	Global (G) or treatment-specific (T)	Estimation method
IGRACER components				
Conventional WTP for QALYs	K	F	G	$K = \frac{C}{\omega_C}$
Elasticity of utility with respect to health	ω_H	F	G	Happiness or DCE, as explained in III.D
Severity ratio	R	V	T	Derive from Eq. (27)
†% QoL loss from permanent disability	d^*	F	T	Outcomes studies
†Disability ratio	ψ	F	T	Derive from Appendix equation (45)
One-period discount rate	β	F	G	Economic literature
Post-treatment survival probability	p	V	T	Outcomes studies
Average increase in survival probability	μ_p	V	T	Outcomes studies
Incremental cost of intervention	ΔC	V	T	Outcomes studies
Marginal rate of substitution between LE and QoL	δ	V	T	$\delta = \frac{\rho H_0}{\omega_H R}$; derive ρ from Eqs. (10–12)
Population incidence of disease	ϕ	F	T	Outcomes studies
Average incremental gain in QoL	μ_B	V	T	Outcomes studies
Certainty–equivalence ratio	ϵ	V	T	Derive using Eq. (32)
Underlying preference parameters				
Annual consumption or income	C	F	G	Economic literature
Elasticity of utility with respect to consumption	ω_C	F	G	Economic literature [11, 14, 15]
Relative risk aversion over QoL	r_H^*	F	G	DCE methods; or happiness methods and Eqs. (24, 25)
Relative prudence over QoL	π_H^*	F	G	DCE methods; or happiness methods and Appendix Eq. (69)
†Relative temperance over QoL	τ_H^*	F	G	DCE methods; or happiness methods and Appendix Eq. (70)
Underlying outcomes parameters				
Change in variance of QoL outcomes	$\Delta\sigma_H^2$	V	T	Outcomes studies
Change in skewness of QoL outcomes	$\Delta[\gamma_1\sigma_H^3]$	V	T	Outcomes studies
†Change in kurtosis of QoL outcomes	$\Delta\kappa_H$	V	T	Outcomes studies
%QoL loss from untreated disease	ℓ^*	V	T	Outcomes studies
%QoL loss from treated disease	t^*	V	T	Outcomes studies
†Baseline QoL in excellent health	H_0	V	T	Assume $H_0 = 1$

†Indicates an optional parameter

$$\frac{\Delta C_0 + \sum_{n=1}^{N} \beta^n \Pi_{n-1}\{p_n \Delta C_n\}}{\sum_{n=1}^{N}\left\{\beta^n \Pi_{n-1}\left(\frac{R_n}{R_j}\right)[\mu_{pn}\delta_n + \phi p_n \mu_{Bn}\epsilon_n]\right\}} \le \frac{KR_j\omega_H}{H_0(1-\psi d^*)}.$$

(35)

As daunting as this might appear, it is simply bookkeeping to summarize results when relevant parameters in the NMB formula vary across multiple periods. In words, where RASA-WTP represents the risk-adjusted and severity-adjusted WTP, Eqs. (34, 35) simply state that:

$$\frac{\text{Discounted incremental costs}}{\text{Discounted generalized risk-adjusted QALY gains}}$$
$$\le \text{RASA-WTP}.$$

(36)

Equations (34, 35) simplify when treatments have stationary distributions of QoL outcomes (the same in every period). In other words, assume that all parameters in Sect. "Guide to using the GRACE method"B remain constant after treatment is administered. Since the survival probabilities are also constant, $\mu_P = \mu_{pn}$, $p_n = p$, and $\Pi_n = p^n$, $\forall n \ge 1$. All other parameters are estimated as explained in Sect. "Guide to using the GRACE method" A, B.

In this case, the *ex ante* net monetary benefit in terms of period zero dollars is:

$$\text{NMB}_{\text{exante}} = \left\{ \beta \left(\frac{1 - (\beta p)^N}{1 - \beta p} \right) \right\}$$
$$\left\{ [\frac{K\omega_H R}{H_0(1 - \psi d^*)}][\mu_p \delta + \phi p \mu_B \epsilon] \right\} \quad (37)$$
$$- \left\{ \Delta C_0 + \sum_{n=1}^{N} (\beta p)^n \Delta C_n \right\}.$$

The first term in curly braces is the discounted N-period time horizon. The second "curly braces" expression is the risk-, severity-, and disability-adjusted value of per-period improvement in life expectancy and QoL, over the relevant time horizon. The final terms measure the discounted cost of treatment over the N-period time horizon.

Turning from net monetary benefit to the corresponding IGRACER, technologies with stationary QoL distributions improve *ex ante* welfare if:

$$\frac{\Delta C_0 + \sum_{n=1}^{N} (\beta p)^n \left\{ \Delta C_n \right\}}{[\mu_p \delta + \phi p \mu_B \epsilon] \beta \left(\frac{1 - (\beta p)^N}{1 - \beta p} \right)} \leq \frac{K\omega_H R}{H_0(1 - \psi d^*)}. \quad (38)$$

which, in words, is identical to Eq. (36).

D. Summary

Applying GRACE requires new estimates of attitudes towards risk in health outcomes (r_H^* and π_H^*), estimation of the rate at which marginal utility in health declines as H increases (ω_C), and estimates of untreated health loss (ℓ^*) associated with specific individual illnesses. These are "done once" parameters that HTA practitioners themselves will not need to develop; they will appear in the health economics and outcomes research literature, estimated by analysts accomplished at such tasks.

Practitioners assessing new medical technologies must accumulate new statistical evidence, including not only mean differences between new treatments and their comparator therapies, but also differences in variance and skewness between treatments and comparators. They must also estimate mean health losses t^* for treated patients, treatment-specific parameters, and measures of uncertainty around those means. GRACE can be applied to static and dynamic contexts, making it suitable for use in Monte Carlo simulations of long-term treatment effects and other similar scenarios. Table 2 summarizes the parameters needed for calculated IGRACERs, along with our proposed estimation methods for each.

As with any value assessment model based on utility-maximization of representative individuals, GRACE cannot fully illuminate concerns about inequities of various kinds, unmeasured externalities like scientific spillovers, or macroeconomic effects arising from major contagious pandemics.

However, combined with a Rawlsian "veil of ignorance" philosophy [22], the GRACE approach might shed light on some of these issues. GRACE highlights the importance (to utility-maximizing individuals) of valuing health improvements more when health losses are greater, either from acute disease or disability. GRACE also highlights how income or baseline health status might affect value assessments, because ω_C, ω_H, and relative risk preferences may vary with income. Combining these insights about "representative individuals" with a "veil of ignorance" philosophy might help guide public policy on these issues. We leave this for others to pursue.

Conclusion

GRACE better aligns value assessment with the preferences of real human consumers. Yet, several complexities in the original development of GRACE could impede its adoption and implementation. This paper provides new analyses that (a) generalize the original GRACE formulation to accommodate permanent disability in the baseline period, (b) generalize the original formulation of GRACE to encompass wider ranges of risk preferences, (c) simplify estimation of necessary parameters, (d) outline tractable methods to estimate necessary parameters to implement GRACE models, and (e) extend GRACE to multi-period contexts. We hope that these refinements, combined with the "guidebook" in Sect. "Guide to using the GRACE method", will help propel forward the implementation of GRACE as the standard HTA model.

The generalization that incorporates disability into the overall WTP value has key public policy importance. The Affordable Care Act prohibits Medicare from using any CEA methods that discriminate against disabled people in making coverage determinations, and similarly prohibits the Patient Centered Outcome Research Institute (PCORI) from using discriminatory CEA methods to measure the value of various medical interventions. The newly refined GRACE model demonstrates that life-extension can be equally or even more valuable to people with disabilities, potentially enabling the use of cost-effectiveness analysis even under current US law.

Finally, our analysis suggests the need for future research on inefficiencies and inequities created by traditional CEA for health insurance programs. Health insurance involves an exchange between consumption today—in the form of premiums paid to private insurers or taxes paid to public insurers—and access to medical technology when sick. In the real-world, health insurance premia and taxes are paid by people of varying health status, often pooled into the same health insurance group. If people with disabilities represent the minority, the presence of healthy consumers serve as

a negative externality, inefficiently reducing the generosity of insurance coverage. To see why, imagine two societies, one in which all insurance participants are perfectly healthy *ex ante* ($d^* = 0$) and one in which they are all disabled *ex ante* ($d^* > 0$). Our analysis implies that optimal insurance coverage could be more generous in the society of the disabled and conversely; this is always true for QoL-improving technologies and may always be true for LE-improving technologies. Since traditional CEA sets coverage decisions as if $d^* = 0$, the resulting coverage may be inefficiently low for disabled people forced to pool with many healthier peers. It thus comes as no surprise that people with disabilities often vigorously oppose the use of traditional CEA to set coverage decisions. Future research should investigate the efficiency rationale for alternative policy solutions, perhaps including the relatively common practice of providing subsidized supplemental coverage for people with disabilities.

As a related matter, future research should estimate how disability affects QoL and QoL utility in healthy and sick states. If disability reduces utility proportionally more (or the same) in healthy states, GRACE implies that disabled people have similar or greater WTP for LE gains. If borne out in the data, this would reinforce GRACE's existing implication of higher WTP for QoL gains among the disabled. GRACE provides a potential path forward for satisfying the requirements of the Affordable Care Act regarding disability-related discrimination, thus allowing use of CEA both in PCORI evaluations and in setting coverage determinations in Medicare.

Appendix I: extensions of GRACE

A. Derivation of disability-adjusted WTP

In the text, we claim that $\frac{W(H_{0d})}{W(H_0)} = 1 - d^*\psi$, where ψ depends on ω_H, relative risk-aversion over QoL (r_H^*), and relative prudence over QoL (π_H^*). We now define ψ more specifically and prove this claim. Begin by computing a Taylor Series expansion of $W(H_{0d})$ around $W(H_0)$:

$$W(H_{0d}) \approx W(H_0) + W'(H_0)(H_{0d} - H_0) + \frac{1}{2}W''(H_0)(H_{0d} - H_0)^2 + \frac{1}{6}W'''(H_0)(H_{0d} - H_0)^3 + \dots. \quad (39)$$

We divide (39) by $W(H_0)$ to recover an expression for $\frac{W(H_{0d})}{W(H_0)}$, and we then analyze the resulting expression term by term:

$$\text{First Term} : \frac{W(H_0)}{W(H_0)} = 1. \quad (40)$$

The second term is $\frac{W'(H_0)(H_{0d}-H_0)}{W(H_0)}$. Multiply and divide this term by H_0, recognizing that the component parts are ω_H and $d^* \equiv \frac{(H_0-H_{0d})}{H_0}$, the expected relative loss in QoL due to permanent disability. With these definitions:

$$\text{Second Term} : -\omega_H d^*. \quad (41)$$

The third term is $\frac{1}{2}\frac{W''(H_0)(H_{0d}-H_0)^2}{W(H_0)}$. Now multiply and divide by $W'(H_0)$. Since absolute risk aversion is $r_H \equiv -\frac{W''(H_0)}{W'(H_0)}$, this becomes $-\frac{1}{2}\{r_H\}\frac{\{E(H_{0d}-H_0)^2\}W'(H_0)}{W(H_0)}$. Now multiply and divide by H_0^2, allocating one H_0 term in the numerator to convert absolute risk aversion, r, into relative risk aversion, r_H^*, and the other to complete ω_H:

$$\text{Third Term} : -\frac{1}{2}\omega_H r_H^* d^{*2}. \quad (42)$$

Note that r_H^* is defined at H_0, whereas, in [1], it is defined at the average level of untreated health, μ_H. Therefore, analysts must either assume that r_H^* is constant across all levels of health (CRRA) or they must estimate r_H^* at baseline health level H_0 as well as at H_{1S}.

The fourth term of (39) is $\frac{1}{6}W'''(H_0)(H_{0d} - H_0)^3$. Dividing by $W(H_0)$, gives $\frac{1}{6}[W'''(H_0)(H_{0d} - H_0)^3]\frac{1}{W(H_0)}$. Defining absolute prudence as $\pi_H \equiv \frac{-W'''(H_0)}{W''(H_0)}$, multiplying and dividing by $W''(H_0)$ gives $-\frac{1}{6}\pi_H[W''(H_0)](H_{0d} - H_0)^3]\frac{1}{W(H_0)}$. Multiplying and dividing by $W'(H_0)$, gives $\frac{1}{6}\pi_H r_H(H_{0d} - H_0)^3]\frac{W'(H_0)}{W(H_0)}$. Next, multiply and divide by H_0^3, allocating two of the H_0 terms in the numerator to the risk-valuation terms and the other to complete the definition of ω_H, Then:

$$\text{Fourth Term} : -\frac{1}{6}\pi_H^* r_H^* \omega_H d^{*3}. \quad (43)$$

As with relative risk-aversion, analysts must either assume that π_H^* is constant across all levels of health (CRRA) or they must estimate π_H^* at baseline health level H_{1S}. Any fifth term involving kurtosis would follow the same general strategy.

In summation, the Taylor Series approximation to $\frac{W(H_{0d})}{W(H_o)}$ is:

$$\frac{W(H_{0d})}{W(H_0)} = \left[1 - d^*(\omega_H + \frac{1}{2}\omega_H r_H^* d^* + \frac{1}{6}\omega_H r_H^* \pi_H^*(d^*)^2 + \dots)\right]. \quad (44)$$

Finally, defining:

$$\psi \equiv \omega_H\left(1 + \frac{1}{2}r_H^* d^* + \frac{1}{6}r_H^* \pi_H^*(d^*)^2 + \dots\right), \quad (45)$$

completes the proof that $\frac{W(H_{0d})}{W(H_0)} = 1 - d^*\psi$.

Inspection of (45) shows that $\psi > 0$, since $\omega_H > 0$ and all other components are weakly positive. Next, we show that $\psi \leq 1$ under weak risk-aversion. Exploiting the definition of d^*, we can rearrange $\frac{W(H_{0d})}{W(H_0)} = 1 - d^*\psi$ to show that:

$$\psi = \frac{\frac{(W(H_0)-W(H_{0d}))}{(H_0-H_{0d})}}{\frac{W(H_0)-W(0)}{H_0-0}}. \tag{46}$$

Here, we implicitly assume that $W(0) = 0$. This equation expresses ψ as the ratio of two slopes: (1) The average slope of W between $H_{0d} > 0$ and H_0; and (2) the average slope of W between 0 and H_0. For a strictly concave, increasing utility function, W, the slope between H_{0d} and H_0 will be strictly less than the slope between 0 and H_0. For a linear utility function, on the other hand, the slopes will be equal. This proves the necessary result.

B. Proof of the reformulation of δ

Define $V \equiv E\big[\phi W(H_{1S} + B)\big] + (1 - \phi)W(H_{1W})$, the expected utility in period 1, conditional on survival to that period. In LP [1], the marginal rate of substitution is defined as $\delta = \frac{V}{W'(\mu_H)}$, the ratio of marginal utility of increased LE to marginal utility of QoL in period 1. In [1], it was necessary to estimate this for each illness/treatment combination.

We begin by proving the equivalence of Eqs. (1) and (9) in the main text. By inspection, we can see that this equivalence rests on the proposition that $\omega_H R\delta = \rho H_0$. To prove the latter, note that LP define $\delta \equiv [\frac{V}{W'(\mu_H)}]$, $R \equiv [\frac{W'(\mu_H)}{W'(H_0)}]$, and $\omega_H \equiv [\frac{W'(H_0)H_0}{W(H_0)}]$. Therefore, $\omega_H R\delta = \left[\frac{W'(H_0)H_0}{W(H_0)}\right]\left[\frac{W'(\mu_H)}{W'(H_0)}\right][\frac{V}{W'(\mu_H)}] = \frac{V}{W(H_0)}H_0$. Since $\rho \equiv \frac{V}{W(H_0)}$, the result follows, and Eqs. (1) and (9) are equivalent.

Next, for compactness, define $H_T \equiv (H_{1S} + B)$, where T stands for "treated" health outcome. Next, define $V_S \equiv W(H_T)$, with a corresponding value of $\rho_S = \frac{V_S}{W(H_o)}$. In parallel, define $V_H = W(H_{1W})$ and $\rho_W = \frac{V_W}{W(H_0)}$. This allows us to think of V as the probability-weighted average of period 1 utility in the sick state and period 1 utility in the well state. It also allows us to rewrite ρ as $\rho = \phi\rho_S + (1 - \phi)\rho_W$.

To estimate ρ_S, consider the Taylor series expansion of V_S around H_0, which proceeds quite similarly to the expansion of $W(H_{0d})$ around $W(H_0)$ above:

Divide (47) by $W(H_0)$, following the definition of ρ_S, and analyze it term by term:

$$First\ Term: \frac{W(H_0)}{W(H_0)} = 1. \tag{48}$$

The second term is $\frac{W'(H_0)E(H_T-H_0)}{W(H_0)}$. Multiply and divide this term by H_0, recognizing that the component parts are ω_H (defined at H_0) and $t^* \equiv \frac{H_0-E(H_T)}{H_0}$, the expected relative loss in health status after treatment. With these definitions:

$$Second\ Term: -\omega_H t^*. \tag{49}$$

The third term is $\frac{1}{2}\frac{W''(H_0)E(H_T-H_0)^2}{W(H_0)}$. Multiply and divide by $W'(H_0)H_0^2$, allocating one H_0 term to the numerator to convert absolute risk aversion, r, into relative risk aversion, r_H^*, and the other to complete ω_H. With these algebraic changes:

$$Third\ Term: -\frac{1}{2}\omega_H r_H^* t^{*2}. \tag{50}$$

As above in Appendix IA, r_H^* is measured at H_0.

The fourth term of (47) is $\frac{1}{6}W'''(H_0)E(H_T - H_0)^3$. Dividing by $W(H_0)$, gives $\frac{1}{6}[W'''(H_0)E(H_T - H_0)^3]\frac{1}{W(H_0)}$. Multiply and divide by $W'(H_0)W''(H_0)H_0^3$. Then:

$$Fourth\ Term: -\frac{1}{6}\pi_H^* r_H^* \omega_H t^{*3}. \tag{51}$$

As before, π_H^* is defined at H_0. Any fifth term involving kurtosis would follow the same general strategy.

In summation, the Taylor Series approximation to $\frac{V_S}{W(H_o)} \equiv \rho_S$ is:

$$\rho_S \approx 1 - \omega_H t^*[1 + \frac{1}{2}r_H^* t^* + \frac{1}{6}\pi_H^* r_H^* t^{*2} + \dots]. \tag{52}$$

Turning to the remaining portion of V, we presume that $H_{1W} = H_{0d}$ in the sense that permanent disability (if present) persists across periods. As in the main text, define $d^* \equiv \frac{H_0-H_{0d}}{H_0} = \frac{H_0-H_{1W}}{H_0}$, the relative QoL loss from permanent disability in either period zero or period one. This allows us to write the parallel Taylor expansion:

$$\rho_W = 1 - \omega_H d^*\left[1 + \frac{1}{2}r_H^* d^* + \frac{1}{6}\pi_H^* r_H^* d^{*2} + \dots\right]. \tag{53}$$

$$E(V_S) \approx W(H_0) + W'(H_0)E(H_T - H_0) + \frac{1}{2}W''(H_0)E(H_T - H_0)^2 + \frac{1}{6}W'''(H_0)E(H_T - H_0)^3 + \dots. \tag{47}$$

Then overall:

$$\rho = \phi\rho_S + (1 - \phi)\rho_W. \tag{54}$$

C. Proof of results relating to disability-adjusted WTP

First, we show that permanent disability increases the WTP for improvements in average QoL. Equation (7) implies that the marginal value of an improvement in average QoL is given by $\frac{K\omega_H R}{H_0(1-\psi d^*)}\phi p_1 \epsilon$. Disability increases $\frac{1}{1-\psi d^*}$. Moreover, if utility exhibits CRRA over QoL, all other terms are unaffected by disability. Technically, disability could have complex effects on ω_H, ψ, and ϵ, if relative risk-aversion varies with QoL. In this case, the relationship between disability and QoL needs to be judged empirically.

Turning to life-extension, Eq. (1) implies that the value of life-extension, μ_p, is given by $K\omega_H R\mu_p\delta$. The total disability-adjusted value of life-extension is thus given by $TDVLE \equiv \frac{K\omega_H R\mu_p\delta}{H_0(1-\psi d^*)}$.

Appendix IB showed that $\delta = \frac{\rho H_0}{\omega_H R}$. Using the definition of ρ, this becomes $\delta = \frac{V H_0}{W(H_0)\omega_H R}$. Moreover, by definition, $\frac{W(H_{0d})}{W(H_0)} = 1 - \psi d^*$. Substituting these two terms into $TDVLE$ yields:

$$TDVLE \equiv K\mu_p\frac{V}{W(H_{0d})} \tag{55}$$

Exploiting the definition of $V \equiv E[\phi W(H_{1S} + B)] + (1 - \phi)W(H_{1W})$ from Appendix IB, this becomes:

$$TDVLE \equiv K\mu_p\left[\phi\frac{E[W(H_{1S} + B)]}{W(H_{0d})} + (1 - \phi)\right]. \tag{56}$$

If permanent disability weakly increases $\frac{E(W(H_{1S}+B))}{W(H_{0d})}$, $TDVLE$ weakly rises (holding μ_p fixed), and vice-versa.

Next, we show for illustrative purposes that people with disabilities have equal WTP for LE under CRRA utility and multiplicative disability. To focus on the disability effects, assume that illness and treatment effects are non-stochastic. Assume further that utility belongs to the HARA family. We implement multiplicative disability by assuming that $H_{0d} = H_0(1 - d^*)$ and $H_{1S} + B = H_0(1 - i_t^*)(1 - d^*)$, where $0 < i_t^* < 1$ represents the percentage loss in QoL from acute treated illness and d^* continues to represent the percentage loss in QoL from disability. Define the functions $\omega_H(H_{0d}) \equiv \frac{W'(H_{0d})H_{0d}}{W(H_{0d})}$ and $\omega_H(H_{1S} + B) \equiv \frac{W'(H_{1S}+B)(H_{1S}+B)}{W(H_{1S}+B)}$, the elasticities of utility with respect to QoL evaluated as indicated. Differentiation implies that:

$$\frac{\partial}{\partial d^*}\left[\frac{W(H_{1S} + B)}{W(H_{0d})}\right]$$
$$= \frac{\left[W'(H_{0d})H_0 W(H_{1S} + B) - W'(H_{1S} + B)H_0(1 - i_t^*)W(H_{0d})\right]}{W(H_{0d})^2}. \tag{57}$$

This can be simplified as:

$$\frac{\partial}{\partial d^*}\left[\frac{W(H_{1S} + B)}{W(H_{0d})}\right] = W(H_{0d})W(H_{1S} + B)$$
$$\frac{\left[\frac{W'(H_{0d})H_{0d}}{W(H_{0d})}\frac{H_0}{H_{0d}} - \frac{W'(H_{1S}+B)(H_{1S}+B)}{W(H_{1S}+B)}\frac{H_0(1-i_t^*)}{H_{1S}+B}\right]}{W(H_{0d})^2}. \tag{58}$$

Observe that, by definition, $\frac{H_0}{H_{0d}} = \frac{1}{1-d^*} = \frac{H_0(1-t^*)}{H_{1S}+B}$. And, when utility is CRRA, Eq. (16) implies that the elasticity of utility is constant, so that $\frac{W'(H_{0d})H_{0d}}{W(H_{0d})} = \frac{W'(H_{1S}+B)(H_{1S}+B)}{W(H_{1S}+B)}$. It thus follows that $\frac{\partial}{\partial d^*}\left[\frac{W(H_{1S}+B)}{W(H_{0d})}\right] = 0$.

Finally, we derive sufficient conditions under which permanent disability increases $\frac{E(W(H_{1S}+B))}{W(H_{0d})}$. Consider the Taylor Series expansion of $E(W(H_{1S} + B))$ around H_{0d}. Following the same approach used in Appendix IA and Appendix IB, one can show:

$$\frac{E(W(H_{1S} + B))}{W(H_{0d})} = 1 - \omega_H\left(\frac{t^* - d^*}{1 - d^*}\right)$$
$$- \omega_H r_H^*\left(\frac{t^* - d^*}{1 - d^*}\right)^2 - \omega_H r_H^*\pi_H^*\left(\frac{t^* - d^*}{1 - d^*}\right)^3 - \dots. \tag{59}$$

Note that in this case, the relative risk preference terms are evaluated at H_{0d}. Under CRRA preferences, the relative risk preference terms do not vary with d^*. In this case, it is clear that $\frac{\partial}{\partial d^*}\frac{E(W(H_{1S}+B))}{W(H_{0d})} \geq 0$ if and only if $\frac{\partial}{\partial d^*}\left(\frac{t^*-d^*}{1-d^*}\right) \leq 0$.

$$\frac{\partial}{\partial d^*}\left(\frac{t^* - d^*}{1 - d^*}\right) = \frac{\left(\frac{\partial t^*}{\partial d^*} - 1\right)(1 - d^*) + (t^* - d^*)}{(1 - d^*)^2}. \tag{60}$$

This derivative will be weakly negative if and only if:

$$\frac{\partial t^*}{\partial d^*} \leq \frac{1 - t^*}{1 - d^*} \leq 1. \tag{61}$$

Expression (61) requires that disability lowers QoL in the sick state by less than in the healthy state. For curative therapies, we require $\frac{\partial t^*}{\partial d^*} \leq 1$, which is always true, but for less than perfect therapies, the requirement becomes non-trivial.

A useful alternate formulation of this expression arises from the change of variables, $D \equiv 1 - d^*$ and $T \equiv 1 - t^*$. D and T represent the QoL index in the

healthy and treated sick states, respectively, in the presence of disability. Notice that $\left(\frac{t^* - d^*}{1 - d^*}\right) = \left(\frac{D - T}{D}\right) = 1 - \frac{T}{D}$. Moreover, $\frac{\partial}{\partial D}\left(\frac{D - T}{D}\right) = -\frac{\left(\frac{\partial T}{\partial D}D - T\right)}{D^2}$. Observe that $\frac{\partial}{\partial D}\left(\frac{D - T}{D}\right) \leq 0$ if and only if $\left(\frac{\partial T}{\partial D}\frac{D}{T}\right) \geq 1$. In other words, reducing disability so that healthy state QoL rises by one percent must increase treated sick state QoL by more than one percent.

Appendix II: simplied hyperbolic absolute risk-aversion (HARA) functions

Traditional HARA utility functions are defined as:

$$U(H) = \frac{1 - \gamma}{\gamma}Z^\gamma, \text{ where } Z = \frac{aH}{1 - \gamma} + \eta. \tag{62}$$

Since H is defined on the [0,1] domain, arbitrary rescaling using the parameter a provides no additional generality. Therefore, a convenient simplification defines $a \equiv (1 - \gamma)$. Then, since $Z = H + \eta$, we can write:

$$U'(H) = (1 - \gamma)Z^{\gamma - 1}, \tag{63}$$

$$U''(H) = (\gamma - 1)(1 - \gamma)Z^{\gamma - 2}, \tag{64}$$

from which we have $r_H = \frac{1 - \gamma}{Z}$ and therefore:

$$r_H^* = (1 - \gamma)\left(\frac{H}{Z}\right). \tag{65}$$

Similarly:

$$\omega_H = \frac{U'(H)H}{U(H)} = \gamma\left(\frac{H}{Z}\right) \tag{66}$$

Importantly, since the $\left(\frac{H}{Z}\right)$ values cancel out, dividing (65) by (66) yields:

$$\frac{r_H^*}{\omega_H} = \frac{1 - \gamma}{\gamma}, \tag{67}$$

from which we have:

$$\gamma = \frac{\omega_H}{\omega_H + r_H^*}. \tag{68}$$

Since $U'''(H) = (\gamma - 2)(\gamma - 1)(1 - \gamma)Z^{\gamma - 3}$, similar calculations demonstrate that:

$$\pi_H^* = (2 - \gamma)\left(\frac{H}{Z}\right) = \frac{(2 - \gamma)}{(1 - \gamma)}r_H^*. \tag{69}$$

Similarly:

$$\tau_H^* = (3 - \gamma)\left(\frac{H}{Z}\right) = \left[\frac{3 - \gamma}{(1 - \gamma)}\right]r_H^*. \tag{70}$$

The general form for the jth relative QoL risk preference parameter, ζ_{jH}^* is:

$$\zeta_j^* = \left[\frac{j - \gamma}{1 - \gamma}\right]r_H^*. \tag{71}$$

Moreover, as proven in [15], Eq. (69) allows the recovery of γ from r_H^* and π_H^* as in:

$$\gamma = \frac{\pi_H^* - 2r_H^*}{\pi_H^* - r_H^*}. \tag{72}$$

Appendix III: consequences of non-CRRA utility for estimation of *R*

Recall eqs. (69) to (71) for definitions of higher-order risk parameters. In contrast, for CRRA utility, $\pi_H^* = 1 + r_H^*$, $\tau_H^* = 2 + r_H^*$, and $\zeta_j^* = (j - 1) + r_H^*$. (Observe that $j = 1$ for relative risk-aversion, $j = 2$ for relative prudence, and so on.) Now suppose that the true utility function is within the HARA family but does not satisfy CRRA (i.e., $\eta \neq 0$). Where $r_H^* = 1$ and using the above expressions, we can estimate the relative bias that would follow from incorrectly assuming CRRA utility. Now define $\widehat{\zeta}_j^*$ as the CRRA-based estimate and ζ_j^* as the true value. Therefore:

$$Relative\ Bias = \frac{\widehat{\zeta}_j^*}{\zeta_j^*} - 1 = (1 - \gamma)\frac{1 + \left[\frac{j - 1}{r^*}\right]}{[j - \gamma]} - 1. \tag{73}$$

With $r^* = 1$, relative bias becomes:

$$Relative\ Bias|_{r^* = 1} = \frac{j(1 - \gamma)}{j - \gamma} - 1. \tag{74}$$

If "relative bias" is 10%, CRRA-based estimates will understate higher-order risk parameters by 10%, etc. Table 3 reports these relative bias figures for various values of γ (along the top row) and for relative risk parameters of differing degrees, j. As noted above, these estimates refer to the case where $r^* = 1$. The left-hand column covers the case of $\gamma = 0$, under which CRRA is the correctly specified utility function (Table 3).

The larger is γ, the greater the potential for relative bias when relying on the CRRA assumption. Indeed, if $\gamma \geq 0.5$, the relative bias approaches or exceeds 50%.

Table 3 assumes $r_H^* = 1$. Equation (73) implies that larger relative risk-aversion parameters somewhat mitigate the

Table 3 Relative bias of relative risk preferences when CRRA is incorrectly specified

	γ					
	0.0	0.2	0.4	0.5	0.6	0.8
π^*	0%	-11%	-25%	-33%	-43%	-67%
τ^*	0%	-14%	-31%	-40%	-50%	-73%
ζ_4^*	0%	-16%	-33%	-43%	-53%	-75%
ζ_5^*	0%	-17%	-35%	-44%	-55%	-76%
ζ_6^*	0%	-17%	-36%	-45%	-56%	-77%

relative bias reported in the table, and vice-versa. Furthermore, all the higher-order terms are moot if the relevant health outcomes have normal distributions, since all statistical moments above variance would then equal zero.

Appendix IV: relationships between utility parameters and their estimation using happiness regression coefficients

First, consider $\omega_H = \frac{W'(H)H}{W(H)}$, the elasticity of utility with respect to H. Then, generally:

$$\frac{\partial \omega_H}{\partial H} = \frac{1}{W}\left[W' + [\frac{W''}{W'}]W\prime H\right] - \frac{1}{W^2}\left[HW'W'\right] = \frac{W'}{W}\left[1 - r_H^*\right] - \left[\frac{\omega_H^2}{H}\right]. \quad (75)$$

From this:

$$\epsilon^\omega = 1 - r_H^* - \omega_H, \quad (76)$$

and thus:

$$r_H^* = \left(1 - \omega_H\right) - \epsilon^\omega. \quad (77)$$

Now consider a second-order happiness regression:

$$\ln(Happy) = \beta_1\ln(H) + \frac{1}{2}\beta_2\ln(H)^2 + \beta_3\ln(C) + \frac{1}{2}\beta_4\ln(C)^2 + \beta_5\ln(H)\ln(C) + \epsilon. \quad (78)$$

Suppressing the interaction term to preserve clarity (i.e., setting $\beta_5 = 0$):

$$\frac{\partial\ln(Happy)}{\partial\ln(H)} \equiv \omega_H = \beta_1 + \beta_2\ln(H). \quad (79)$$

This implies the empirical analogs:

$$\widehat{\omega_H} = \hat{\beta}_1 + \hat{\beta}_2\ln(H), \quad (80)$$

$$\widehat{\frac{\partial\omega_H}{\partial H}} = \frac{\hat{\beta}_2}{H}. \quad (81)$$

From this:

$$\widehat{\epsilon^\omega} = \frac{\partial\widehat{\omega_H}}{\partial H}\left(\frac{H}{\widehat{\omega_H}}\right) = \frac{\hat{\beta}_2}{\widehat{\omega_H}}. \quad (82)$$

Finally, using eq. (80) and (82), and referring to (77), we can estimate relative risk-aversion as:

$$\widehat{r_H^*} = \left(1 - \widehat{\omega_H}\right) - \frac{\hat{\beta}_2}{\widehat{\omega_H}} = 1 - \left(\hat{\beta}_1 + \hat{\beta}_2\ln(H)\right) - \frac{\hat{\beta}_2}{\left(\hat{\beta}_1 + \hat{\beta}_2\ln(H)\right)}. \quad (83)$$

Higher-order estimates follow by using (68) to estimate γ using estimates of ω_H and r_H^*, and by using eq. (69) to (71) to compute the higher-order risk preference parameters using estimates of γ and r_H^*.

Declarations

Conflicts of interest DNL discloses that he holds equity in Precision Medicine Group and has received research support, consulting fees, or speaking fees from Amgen, Biogen, Edwards Lifesciences, Genentech, GRAIL, Otsuka, Novartis, Perrigo, and Pfizer. CEP discloses consulting fees from Biogen, Pfizer, Audentes Therapeutics, and the Office of Health Economics (London).

Data availability Not applicable.

Code availability Not applicable.

References

1. Lakdawalla, D., Phelps, C.: Health Technology Assessment with Risk Aversion in Health. J. Health Econ. **72**, 102346 (2020)
2. Lakdawalla, D.N., Phelps, C.E.: Health Technology Assessment with diminishing returns to health: the generalized risk-adjusted cost effectiveness (GRACE) approach. Value Health **24**(2), 244–249 (2021)
3. Garber, A.M., Phelps, C.E.: Economic foundations of cost-effectiveness analysis. J. Health Econ. **16**(1), 1–31 (1997)
4. Attema, A.E., et al.: An elicitation of utility for quality of life under prospect theory. J. Health Econ. **48**, 121–134 (2016)
5. Nielsen, J.S., Gyrd-Hansen, D., Kjær, T.: Sample Restrictions and the elicitation of a constant willingness to pay per quality adjusted life year. Health Econ. **30**(5), 923–931 (2021)
6. Nord, E., et al.: Maximizing health benefits vs egalitarianism: an Australian survey of health issues. Soc. Sci. Med. **41**(10), 1429–1437 (1995)
7. Green, C., Gerard, K.: Exploring the social value of health-care interventions: a stated preference discrete choice experiment. Health Econ. **18**(8), 951–976 (2009)
8. Linley, W.G., Hughes, D.A.: Societal views on nice, cancer drugs fund, and value based pricing criteria for prioritising medicines: a cross-sectional survey of 4118 adults in Great Britain. Health Econ. **22**(8), 948–964 (2013)
9. Pratt, J.W.: Risk aversion in the small and in the large. Econometrica **32**(1), 122–136 (1964)
10. Kimball, M.S.: Precautionary saving in the small and in the large. Econometrica **58**(1), 53–73 (1990)
11. Phelps, C.E.: A new method to determine the optimal willingness to pay in cost-effectiveness analysis. Value Health **22**(7), 785–791 (2019)
12. Nord, E., et al.: Incorporating societal concerns for fairness in numerical valuations of health programmes. Health Econ. **8**(1), 25–39 (1999)
13. Basu, A., Carlson, J., Veenstra, D.: Health years in total: a new health objective function for cost-effectiveness analysis. Value Health **23**(1), 96–103 (2020)
14. Rosen, S.: The value of changes in life expectancy. J. Risk Uncertain. **1**(3), 285–304 (1988)
15. Phelps, C.E., Cinatl, C.: Estimating optimal willingness to pay thresholds for cost-effectiveness analysis: a generalized method. Health Econ. (2021). https://doi.org/10.1002/hec.4268
16. Noussair, C.N., Trautmann, S.T., Van de Kuilen, G.: Higher order risk attitudes, demographics, and financial decisions. Rev. Econ. Stud. **81**(1), 325–355 (2013)
17. Johnson, F.R., et al.: Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. Value Health **16**(1), 3–13 (2013)
18. Easterlin, R.A.: Explaining happiness. Proc. Natl. Acad. Sci. **100**(19), 11176–11183 (2003)
19. Christensen, L.R., Jorgenson, D.W., Lau, L.J.: Transcendental logarithmic utility functions. Am. Econ. Rev. **65**(3), 367–383 (1975)
20. World Health Organization: Who Methods and Data Sources for Global Burden of Disease Estimates 2000–2016. World Health Organization, Geneva (2018)
21. Center for the Evaluation of Value and Risk in Health. The Cost-Effectiveness Analysis Registry [Internet]. Available from: www.cearegistry.org (2021)
22. Rawls, J.A.: A Theory of Justice . Belknap Press, Cambridge (1971)