Journal of Benefit-Cost Analysis (2023), 14: 1, 44–67
doi:10.1017/bca.2023.6

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

The Generalized Risk-Adjusted Cost-Effectiveness (GRACE) Model for Measuring the Value of Gains in Health: An Exact Formulation

Darius N. Lakdawalla[1,*] and Charles E. Phelps[2]

[1]Schaeffer Center for Health Policy and Economics, University of Southern California, Los Angeles, CA, USA
[2]Department of Economics, University of Rochester, Rochester, NY, USA
*Corresponding author: e-mail: dlakdawa@usc.edu

Keywords: expected utility, health technology assessment, marginal utility of health, risk aversion

JEL classification: I1, I13, I18

Abstract

The generalized risk-adjusted cost-effectiveness (GRACE) analysis method modifies standard cost-effectiveness analysis (CEA), the primary method currently used worldwide to value health improvements arising from healthcare interventions. Generalizing standard CEA, GRACE allows for decreasing or even increasing returns to health. Previous presentations of GRACE have relied extensively on Taylor Series expansion methods to specify key model parameters, including those that properly adjust for illness severity and preexisting disability, consequences of uncertain treatment outcomes, and the marginal rate of substitution between life expectancy and health-related quality of life. Standard CEA cannot account for these sources of value or cost in its valuation of medical treatments. However, calculations of GRACE measures based on Taylor Series are approximations, which may be poorly behaved in some contexts. This paper provides a new approach for implementing GRACE, using exact utility functions instead of Taylor Series approximations. While any proper utility function will suffice, we illustrate with three well-known functions: constant relative risk aversion (CRRA) utility; hyperbolic absolute risk aversion (HARA) utility, of which CRRA is a special case; and expo-power (EP) utility, of which constant absolute risk aversion (CARA) is a special case. The analysis then extends from two-period to multiperiod models. We discuss methods to estimate parameters of HARA and EP functions using two different types of data, one from discrete choice experiments and the other from "happiness economics" methods. We conclude with some reflections on how this analysis might affect benefit-cost analysis studies of healthcare interventions.

1. Introduction

Cost-effectiveness analysis (CEA), the primary method to measure the value of healthcare interventions (Garber & Phelps, 1997), is widely used around the world currently as an alternative to benefit-cost analysis (BCA). This emerged in healthcare because CEA practitioners were unwilling to assume a "value of life," that is, a specific value of the

decision threshold, K, the standard CEA willingness to pay (WTP) for gains in health. They instead reported only incremental cost-effectiveness ratios (ICERs) for various healthcare interventions. This leaves the monetary value of the WTP to social decision-makers, who would then determine if an intervention was worthy of use. This tradition began primarily in the British National Health Service after World War II and has since spread around the world.

Unfortunately, CEA imposes an important restriction on the allowed form of utility – that there are constant returns to health in creating utility. This assumption appears at odds with evidence that consumers value treating severe disease more than mild disease. To relax this restriction, the generalized risk-adjusted cost-effectiveness (GRACE) model has been developed (Lakdawalla & Phelps, 2020; Lakdawalla & Phelps, 2021, 2022; Phelps & Lakdawalla, 2023). CEA is a special case of GRACE.

To date, GRACE has been explicated using Taylor Series expansions (Taylor, 1715) to estimate the necessary parameters in the model, thus allowing full flexibility in estimating the GRACE measure of value for medical interventions. However, in practice, convergence of some of the Taylor Series approximations is slow, particularly in situations involving highly severe illness and/or disabilities. Further, the Taylor Series presentations do not always generate an intuitive understanding of the GRACE methodology.

An alternative and perhaps more useful approach assumes that utility in health has a specific functional form, a common approach, for example, in studies of the economic value of extending life expectancy (LE; Cordoba & Ripoll, 2017), the estimation of risk preferences (Holt & Laury, 2002; Noussair et al., 2014; Holt & Laury, 2014) and various other health economics analyses (e.g., Zeckhauser, 1970; Feldstein, 1973; Feldstein & Friedman, 1977; Keeler et al., 1988; Manning & Marquis, 1996; Marquis & Holmer, 1996; Garber & Phelps, 1997). The purpose of this article is to present the GRACE methodology with exact utility functions.

We first summarize the GRACE model and then describe the total value of a medical intervention using constant relative risk aversion (CRRA) utility and the more general hyperbolic absolute risk aversion (HARA) model. Next, we briefly discuss a third alternative, the expo-power (EP) model (Saha, 1993). CRRA utility is a special case of HARA. Constant absolute risk aversion (CARA) is a special case of EP. Since CRRA utility provides the clearest conceptual understanding of how GRACE works, we use that as the primary presentation vehicle, followed by a discussion of the changes that occur when the more general HARA function is used and implications for use of the EP model. A glossary of the acronyms and parameters used herein appears at the end of the article.

2. Overview of CEA and GRACE

2.1. Standard cost-effectiveness and WTP for health gains

The formal version of cost-effectiveness analysis (CEA), consistent with microeconomic principles, was first set forth by Garber and Phelps (1997). They assumed that health is produced with a two-input process, $H = H(a,b)$, where p_a and p_b are the prices of a and b, and H'_a and H'_b are the associated marginal products. Define utility as a function of consumption, C, equal to income less medical spending, and health, H. In a two-period model, H_0 is baseline period zero health, measured on a scale of $0 \leq H \leq 1$, and typically set at $H_0 = 1$,

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

perfect health. Overall utility is given by $V(C,H) = U(C)H$, where $U(C)$ is assumed to have diminishing returns, but utility is assumed to have constant returns to H. Defining the elasticity of utility with respect to C as $\omega_C \equiv U'(C)\left[\frac{C}{U(C)}\right]$, the WTP for a quality-adjusted life-year (QALY) is the marginal utility of a QALY scaled by the marginal utility of consumption, or:

$$K \equiv \frac{U(C)}{U'(C)H_0} = \left[\frac{C}{H_0}\right]\left[\frac{1}{\omega_C}\right]. \tag{1}$$

Insurance coverage choices are made in period 0, before individual illness states are known. These choices affect access to care in period 1, after illness states are realized. Expected utility is maximized by adjusting the use of inputs so that $\frac{p_a}{H'_a} = \frac{p_b}{H'_b} \leq K$. WTP for one QALY exceeds the value of a year's consumption, C, only because of diminishing returns to consumption, that is, $0 < \omega_C < 1$. Current estimates suggest that $2C \leq K \leq 3C$ in industrialized countries, lower in developing countries (Phelps, 2019; Phelps & Cinatl, 2021; Phelps & Lakdawalla, 2023).

Once K is determined, CEA and BCA are equivalent under certain conditions (Bleichrodt & Quiggin, 1999). The CEA rule says that the incremental cost-effectiveness ratio (ICER) of any medical intervention – marginal cost divided by marginal product – should not exceed WTP, normally stated as $\frac{p_a}{H'_a} \leq K$. Once a specific value of K is chosen, CEA value measures convert directly into net monetary benefit (NMB) measures and hence to standard BCA (Phelps & Mushlin, 1991).

Since CEA presumes constant returns to H, it does not matter to whom health improvements are given, what their untreated illness severity is, what their level of preexisting disability is, or what degree of uncertainty about treatment benefits exists. Constant returns to H also require that the marginal rate of substitution (MRS) between gains in health-related quality of life (HRQoL) and LE be the same in all situations. The mantra of standard CEA says that "… a QALY is a QALY is a QALY…" (Williams, 1992). Mathematically, the total value of a medical intervention in CEA is

$$\text{TVMI}_{\text{CEA}} = \text{LE} \times \Delta\text{HRQoL} + \text{HRQoL} \times \Delta\text{LE}. \tag{2}$$

This formula values gains in LE at the existing HRQoL, and thus values LE gains for disabled people less than the value for otherwise-similar nondisabled people. Similarly, if disability reduces LE, then gains in HRQoL are similarly lower for disabled than for nondisabled people. This may have contributed to the U.S. Affordable Care Act's banning of cost-effectiveness measures that discriminate against disabled people for federally related health programs.

2.2. How GRACE differs from CEA

GRACE retains CEA's assumption of separable utility in consumption, C, and health, H[1]:

$$V(C,H) = U(C)W(H). \tag{3}$$

[1] Separability simplifies the presentation without altering any basic concepts in the model. It limits the ways in which H and C interact in creating utility rather than allowing more-general interactions. With separability, $\frac{\partial^2 V(C,H)}{\partial C \partial H} = \frac{\partial^2 V(C,H)}{\partial H \partial C} = U'(C)W'(H)$.

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

However, in GRACE, both $U(C)$ and $W(H)$ exhibit positive but diminishing returns to health.[2] As proven in Lakdawalla and Phelps (2020) and Lakdawalla & Phelps (2021, 2022), this simple generalization alters standard CEA methods in six important ways:

(i) GRACE demonstrates greater WTP for HRQoL gains as untreated illness severity increases.
(ii) GRACE reveals greater WTP for both HRQoL and LE gains as disability increases.
(iii) GRACE shows that with diminishing returns to health, standard CEA over-values gains in health, with the magnitude of over-valuation depending on how rapidly marginal utility declines as health increases.
(iv) Combining these three insights, standard CEA methods overvalue treatments for low-severity illnesses and undervalue treatments for high-severity illnesses, and they undervalue treatments for persons with preexisting disabilities.
(v) The relative value of LE and HRQoL varies by initial values of those measures, contrary to the CEA implication that it is identical in all situations.
(vi) When consumers exhibit relative risk aversion and prudence (Kimball, 1990), uncertainty in treatment outcomes for HRQoL lowers the value of treatment gains, but increases in the probability of unusually good outcomes (positive skewness in HRQoL outcome distributions) increase the value of treatments.

GRACE defines five new parameters that can be computed using knowledge of underlying preference parameters, along with measures of treated and untreated illness and disability, as well as the distributions of health outcomes in the treated and untreated states.

We also note that, like traditional CEA and almost the entire health economics literature, GRACE presumes time-separable utility, which in turn implies consumer risk-neutrality over changes in mortality risk. Therefore, GRACE primarily affects the value of HRQoL, rather than of LE gains.

2.2.1. Nonstochastic parameters

Three GRACE parameters depend only on levels of health outcomes. These parameters combine to alter the traditional CEA value per unit of health gain, K, to reflect the consequences of diminishing return to health. We discuss these next.

Diminishing returns to H. From Equation (3),

$$\frac{\partial V(C,H)}{\partial C} = U'(C)W(H), \tag{4a}$$

$$\frac{\partial V(C,H)}{\partial H} = W'(H)U(C). \tag{4b}$$

Therefore, the MRS between C and H is

$$\text{MRS} \equiv \frac{W'(H_0)U(C_0)}{U'(C_0)W(H_0)}. \tag{4c}$$

[2] GRACE also allows for regions of strict convexity in W, if, for example, patients exhibit "value of hope" (Lakdawalla et al., 2012). We focus on the globally concave case for expositional convenience.

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

Define $\omega_H = W'(H_0)[\frac{H_0}{W(H_0)}]$, the elasticity of utility with respect to health. Treat baseline health as "perfect" so that $H_0 = 1$, and define baseline consumption, C_0. Then,

$$\text{MRS} = \left[\frac{C_0}{H_0}\right]\left[\frac{\omega_H}{\omega_C}\right]. \tag{4d}$$

This alters the WTP value K from Equation (1) in several distinct ways. Using Equation (4d), we define the first adjustment to $K = \left[\frac{C_0}{\omega_C}\right]\left[\frac{1}{H_0}\right]$ as

$$K^* = K\omega_H. \tag{5a}$$

Since $0 < \omega_H \leq 1$, $K^* \leq K$. Therefore, before adjusting for illness severity or preexisting disability, CEA weakly overvalues gains in HRQoL compared with GRACE, with equality obtaining only under the standard CEA restriction that $\omega_H = 1$.

Illness severity. The second GRACE change introduces illness severity. Define ℓ^* as the proportional loss in untreated health in period 1. Further, define $H_{1S} = H_0(1-\ell^*)$ as the untreated health level in period 1 after the illness occurs and $\mu_H \equiv E(H_{1S})$ as its mean. For ease of intuition, GRACE uses mean QoL gain as the unit of health improvement; later, we explain how GRACE accounts for variance and higher-order moments of QoL gain. As with standard CEA, GRACE assumes full annuitization and constant consumption, so that $C_0 = C_1 = C$. Therefore, $\text{MRS} = \frac{W'(\mu_H)U(C)}{U'(C)W(H_0)}$. In effect, the MRS in Equation (4c) is multiplied by $R = \frac{W'(\mu_H)}{W'(H_0)}$, the ratio of marginal utilities in the average untreated sick state to the healthy state. This adjustment shifts the location at which MRS is measured on indifference curves from H_0 to μ_H. Lakdawalla and Phelps (2020) prove that the WTP measure K in standard CEA then becomes

$$K_{\text{GRACE}} = RK^* = K\omega_H R. \tag{5b}$$

For low-severity illnesses, $R \approx 1$, and R rises exponentially with severity. Since $0 < \omega_H < 1$ under diminishing returns, it follows that standard CEA overvalues treatments for low-severity illnesses and undervalues them for high-severity illnesses.

Permanent disability. Lakdawalla and Phelps (2021, 2022) further generalized this approach by introducing the possibility of preexisting permanent disability. Just as the MRS between well and sick states changes with untreated illness severity, so also the MRS changes with permanent disability, rising as the degree of disability increases. In perfect health, before any illness has presented itself, the period zero $\text{MRS} = \left[\frac{C}{H_0}\right]\left[\frac{\omega_H}{\omega_C}\right]$. Define d^* as the proportional loss in HRQoL created by disability and $H_D = H_0(1-d^*)$. To properly represent the ratio of marginal utilities when preexisting disability is present (see Equation (4c)), the MRS must be adjusted by the factor

$$D = \frac{W(H_0)}{W(H_D)}. \tag{5c}$$

$D = 1$ with no disability, and D rises exponentially as d^* rises. This adjustment is necessary to correct for the new base from which *ex ante* resource allocation decisions are made, accounting for the way the utility of health, $W(H)$, affects the marginal utility of consumption in Equation (4a) and hence in the MRS in Equation (4c).

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

The GRACE-adjusted WTP. These changes lead to the final difference in WTP compared with standard CEA:

$$K^{D}_{\text{GRACE}} = DK_{\text{GRACE}} = DRK^{*} = DR\omega_{H}K. \tag{6}$$

CEA is a special case of GRACE, characterized by the restrictions that $d^{*} = 0$ and $\omega_{H} = 1$ so that R and D and H_0 all equal 1.0, and WTP collapses to the standard CEA measure of K. GRACE's new parameters rely only on parameters of utility functions themselves and upon levels of HRQoL in treated and untreated states. Therefore, estimating them does not require information about the distribution of treatment outcomes.

A graphical presentation of these changes may further assist in understanding how GRACE differs from standard CEA. These figures show indifference curves between consumption, C and health, H. The MRS is the negated slope of the indifference curve at any point along it, and it equals WTP for health gains in each condition.

Setting aside, for now, any randomness in H_{1S}, Figure 1 shows how the MRS increases when moving from perfect health, H_0, to untreated health with an illness in period 1, $H_{1S} = H_0(1 - \ell^{*})$. The second panel in Figure 1 shows the additional increase in the MRS for a person who begins with a permanent disability and then incurs the same illness. In this case, the untreated level of illness is $H_{1S} = H_0(1 - \ell^{*})(1 - d^{*})$.This demonstrates items (i) and (ii) shown above.

Figure 2 shows how the MRS changes as the elasticity of utility with respect to consumption, ω_C changes. In general, $MRS = \frac{\omega_H}{\omega_C}\frac{W(H_{1S})}{W(H_0)}\frac{C}{H_{1S}}$. Under the assumptions of standard CEA, this collapses to $K = \frac{C}{\omega_C}$. As ω_C declines from ω^{1}_{C} to ω^{2}_{C} and ω^{3}_{C}, the opportunity cost of consumption falls, and the WTP, the MRS in this figure, grows as the indifference curves steepen.

In each of these curves, the elasticity of utility with respect to health, ω_H, is kept at 1.0, the standard CEA assumption. Thus, Figure 2 demonstrates how the CEA measure of WTP, K, changes as ω_C changes.

Figure 3 shows the effect of introducing the fundamental GRACE change in assumptions about the nature of utility. For any one of the curves shown in Figure 2, if the elasticity of utility with respect to health, ω_H, embeds positive but strictly diminishing returns, that is, $0 < \omega_H < 1$, then the indifference curves flatten out, so the MRS falls at every point on the curve. Figure 3 shows the primary effect of allowing diminishing returns, item (iii) shown above.

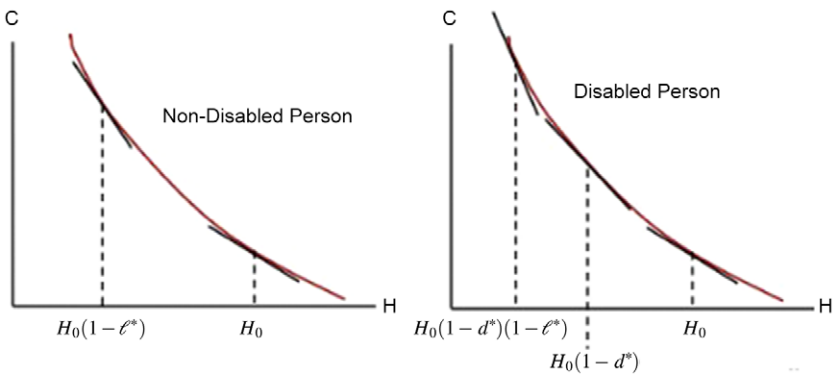

Figure 1. *MRS between consumption and HRQoL increases with acute illness severity and/or permanent disability.*

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

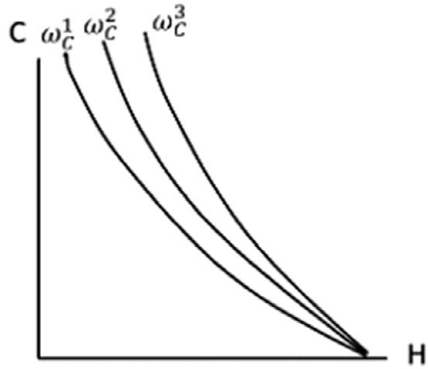

Figure 2. *The willingness to pay for health improvement at any given illness severity rises as the utility elasticity of consumption falls.*

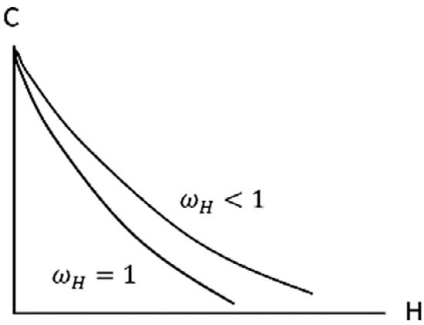

Figure 3. *The willingness to pay for health improvement at any given illness severity rises as the utility elasticity of health falls.*

In combination, these results demonstrate how GRACE changes WTP from the standard CEA value of K to $K^D_{\text{GRACE}} = DR\omega_H K$.

2.2.2. Stochastic parameters

The previous discussion dealt with three GRACE parameters that do not involve the distribution of health outcomes. Next, we discuss two additional parameters that require information about those distributions.

Uncertainty of treatment outcome. Lakdawalla and Phelps (2020, Equation (5)) define the expected period utility gain of a stochastic treatment B for a patient known to have the disease in question as

$$EW(T) = E[W(H_{1S} + B) - W(H_{1s})] \tag{7a}$$

and the expected monetary value of that treatment gain as[3]

[3] From Lakdawalla and Phelps (2020, Equation (5)), this sets the probabilities of survival into period 1, p_1 and the probability of illness, ϕ equal to 1 to minimize notational clutter.

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

$$EV(B) = \left[\frac{U(C)}{U'(C)}\right]\left[\frac{EW(T)}{W(H_0)}\right]. \tag{7b}$$

Next, define

$$ETG = \frac{EW(T)}{W'(\mu_H)}. \tag{7c}$$

Normalizing $EW(T)$ by $W'(\mu_H)$ converts it into units of average HRQoL improvement, ETG.

Recalling that $R \equiv \frac{W'(\mu_H)}{W'(H_0)}$, simple algebraic manipulation now gives

$$EV(B) = K\omega_H R\{ETG\}. \tag{7d}$$

By computing a Taylor series expansion of $EW(T)$ around the level of untreated health, Lakdawalla and Phelps (2020) show that $\mu_B\epsilon \approx ETG$, where μ_B is the average HRQoL improvement, and ϵ is the "certainty-equivalence" ratio measuring the number of riskless QoL units the individual would give up in exchange for the treatment's stochastic HRQoL gains. The certainty-equivalence ratio, ϵ, allows for the approximation of health gains in terms of nonstochastic QoL improvement. It can be computed using the information on higher-order parameters of the statistical distributions of treated and untreated outcomes, including mean, variance, skewness, and kurtosis. We shortly use this Taylor Series approximation in a comparison of GRACE and standard CEA methodologies.

The MRS between LE and HRQoL. To convert longevity and HRQoL gains into a single index of health improvement, GRACE, like traditional CEA, converts longevity gains into units of HRQoL improvement. This requires use of the MRS between LE gains and average HRQoL gains. Furthermore, GRACE, like traditional CEA and almost all the health economics literature, assumes that utility is time-separable and thus linear in the probability of survival, holding consumption, and HRQoL constant. Therefore, the marginal utility of gains in LE is the expected utility of the level of health experienced in after the effects of treatment on HRQoL are received: $E(W(H_T))$. The marginal utility of gains in average HRQoL is simply $W'(\mu_H)$, the marginal utility of the average value of untreated health. The ratio of these forms the MRS between LE and HRQoL, δ, and has units of measurement of H

$$\delta = \left\{\frac{E[W(H_T)]}{W'(\mu_H)}\right\}. \tag{8a}$$

Lakdawalla and Phelps (2022) define a new variable, the ratio of expected utility in the treated state to utility of perfect health,

$$\rho = \frac{E[W(H_T)]}{W(H_0)}. \tag{8b}$$

They then prove that

$$\omega_H R\delta = \rho H_0. \tag{8c}$$

Therefore, $K\omega_H R\delta\mu_P = K\rho H_0\mu_P$. We will use this alternative expression below.

Under traditional CEA, $E(W(H_T)) = H_T$, $W'(\mu_H) = 1$, and the MRS becomes simply H_T. However, with diminishing returns to H, people should be more willing to trade LE for HRQoL gains when HRQoL is low, and conversely. Put slightly differently, $\frac{1}{\delta}$ represents the WTP for HRQoL gains expressed in terms of LE, not consumption.

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

2.2.3. The total value of medical interventions in GRACE

Combining these parameters, we are now in a position to specify the total value of medical interventions (TVMI) for GRACE. Define the mean gain in survival probability (LE gain) from treatment as μ_P, and define the *ex ante* probability of illness as ϕ. Now, from the *ex ante* period 0 perspective,

$$TVMI_{GRACE} = KD\phi\{\mu_P \rho H_0 + \mu_B \epsilon \omega_H R\}. \tag{9a}$$

In contrast and relying on Equation (2), traditional CEA computes TVMI as

$$TVMI_{CEA} = K\phi[\mu_P H_T + \mu_B]. \tag{9b}$$

This highlights the differences between GRACE and CEA. In CEA where $\omega_H = 1$, then so also $R = 1, \epsilon = 1, \rho = H_T$ and $D = 1$. Compared with GRACE, CEA overlooks the effects of the effect of diminishing returns to health, ω_H, the extra value of treating disabled people, D, the effect of illness severity on treatment, R, the effect of QoL levels on the MRS between LE and HRQoL, ρ, and the effect of treatment uncertainty on value, ϵ.

The terms in Equation (9a) can be recovered without knowledge of a specific utility function, provided the analyst can estimate relative risk aversion, relative prudence, and other higher-order risk preferences (Lakdawalla & Phelps, 2020; Lakdawalla & Phelps, 2022). However, this process introduces approximation error into the TVMI expression, because of the Taylor series expansions used to estimate the parameters in the equation. As an alternative, we can unravel the definitions of $D, \rho, \epsilon, \omega_H$, and R to express an "exact" form of Equation (9a):

$$TVMI_{GRACE} = \frac{K\phi H_0}{W(H_0(1-d^*))}\{\mu_P E[W(H_T)] + E[W(T)]\}. \tag{9c}$$

The first term in curly braces describes the expected utility from gains in LE and the second term describes the expected utility from gains in HRQoL, as defined in Equation (7a). When the function W is known, along with the distributions of H_{1s} and B, all terms in Equation (9c) can be calculated exactly, without reliance on approximations.

2.2.4. Summary

In all previous explications of GRACE, each key parameter has been estimated using Taylor series expansion methods. In what follows, we replace the Taylor Series expansion methods with the use of exact utility functions. This simplifies the understanding of the GRACE methods and the estimation of the $TVMI_{GRACE}$.

3. GRACE estimation with exact utility functions

This section develops the entire GRACE model using specific utility functions. In concept, this same approach could be used for any specified utility function. Since computing Equation (9c) for any given utility function is simply an algebraic exercise, we do not labor through that process here. Rather, we explore what several different utility functions imply for key GRACE parameters to provide intuition.

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

3.1. Constant relative risk aversion

The simplest case to understand is to specify that utility has CRRA. This formulation lends itself to intuitive understanding of the way GRACE functions. With CRRA, utility and marginal utility are given by

$$W(H) = \left[\frac{1-\gamma}{\gamma}\right] H^{\gamma} \tag{10a}$$

and

$$W'(H) = (1-\gamma)H^{\gamma-1}. \tag{10b}$$

With CRRA utility, relative risk aversion is given by $r_H^* = (1-\gamma)$, and the elasticity of utility with respect to health is $\omega_H = \gamma$. Further, $r_H^* + \omega_H = 1$. This simplified structure allows us to easily calculate exact forms for the key GRACE parameters. Positive but diminishing marginal utility requires that $0 < \gamma < 1$.

3.1.1. Nonstochastic parameters

First, as noted, with CRRA, $\omega_H = \gamma$. Now, consider R, the severity of illness multiplier, and D, the adjustment to WTP for permanent disability. Since $H_{1S} = H_0(1-\ell^*)$, using Equation (10a), we can readily write R, the ratio of two marginal utilities, as

$$R = \left[\frac{H_{1S}}{H_0}\right]^{\gamma-1} = \left[\frac{H_0}{H_0(1-\ell^*)}\right]^{1-\gamma} = \left[\frac{1}{1-\ell^*}\right]^{1-\gamma} = \left[\frac{1}{1-\ell^*}\right]^{r_H^*}. \tag{11a}$$

When $r_H^* = 0$, the standard CEA assumption, then $R = 1$.

Next, D is a ratio of utilities, not marginal utilities, so where $H_{1d} = H_0(1-d^*)$, using Equation (10a)

$$D = \left[\frac{H_0}{H_0(1-d^*)}\right]^{\gamma} = \left(\frac{1}{1-d^*}\right)^{\gamma} = \left(\frac{1}{1-d^*}\right)^{\omega_H}. \tag{11b}$$

Therefore, $D \geq 1$ and increases exponentially with the severity of disability, d^*. When $d^* = 0$, the standard CEA assumption, then $D = 1$.

Since $\omega_H = \gamma$ with CRRA utility, we can now directly compute the multiplier that adjusts for diminishing returns to H. Combining these, the GRACE value per unit of health gain is

$$K_{GRACE}^{D} = K\omega_H DR = K\gamma\left[\frac{1}{1-d^*}\right]^{\gamma}\left[\frac{1}{1-\ell^*}\right]^{1-\gamma}. \tag{12}$$

Asymptotically, CRRA utility is linear in H when $\gamma = 1$, that is, standard CEA. In this case, adopting the usual CEA perspective that $d^* = 0$, the value per unit of health gain collapses to K.

3.1.2 Stochastic parameters

The MRS between LE and HRQoL. The general equation for this tradeoff is given in Equation (8a). To fully specify this, we must first define the probability distribution of outcomes for individuals who are sick but treated. Define each possible treated outcome as H_n^T for $n = 1 \ldots N$, each with an associated probability π_n^T. Then, using Equation (8b),

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

$$\rho H_0 = H_0\left[\sum\nolimits_n \pi_n^T\left(\frac{W\left(H_n^T\right)}{W(H_0)}\right)\right]. \tag{13a}$$

When utility is CRRA

$$\rho H_0 = H_0\sum\nolimits_n \pi_n^T\left(\frac{W\left(H_n^T\right)}{W(H_0)}\right) = H_0\sum\nolimits_n \pi_n^T\left(\frac{H_n^T}{H_0}\right)^{\gamma}. \tag{13b}$$

Defining $H_n^T = H_0\left(1 - t_n^*\right)$, then

$$\rho H_0 = H_0\sum\nolimits_n \pi_n^T\left(\frac{H_0(1-t_n^*)}{H_0}\right)^{\gamma} = H_0\sum\nolimits_n \pi_n^T\left((1-t_n^*)^{\gamma}\right). \tag{13c}$$

Note that when constant returns to H are specified, then, asymptotically, $\gamma = 1$ and $H_0\sum_n \pi_n^T\left(\frac{H_n^T}{H_0}\right)^{\gamma}$ collapses to $\sum_n \pi_n^T\left(H_n^T\right)$, so ρH_0 equals expected treated health in period 1, identified in Garber and Phelps as H_1. GRACE differs from CEA in this calculation, using CRRA utility, by the effect of the power exponent γ.

Uncertain treatment outcomes. Adding the disability adjustment to Equation (7b), and recalling that $K = \left[\frac{C}{\omega_C}\right]\left[\frac{1}{H_0}\right]$, the value of the medical intervention (in units of consumption, e.g., dollars) is

$$EV(B) = K\omega_H RD\{ETG\}. \tag{14a}$$

To complete Equation (14a) using CRRA, define the possible treatment outcomes as H_{T_i} and their associated probabilities, π_i^T. Similarly, define the possible outcomes in the untreated state as H_{1S_j} and their associated probabilities as π_j^U. Then, *ETG*, measured in units of HRQoL gains, becomes

$$ETG = \left\{\sum\nolimits_i \pi_i^T \frac{W(H_{T_i})}{W'(H_0)} - \sum\nolimits_j \pi_j^U \frac{W\left(H_{1S_j}\right)}{W'(H_0)}\right\}. \tag{14b}$$

Finally, applying the CRRA definitions of $W(H)$ and $W'(H)$ from Equations (10a) and (10b), we have

$$EV(B) = K\omega_H RD\left\{\sum \pi_i^T\left[\frac{H_0}{\gamma}\right]\left(\frac{H_{Ti}}{H_0}\right)^{\gamma} - \left[\frac{H_0}{\gamma}\right]\sum \pi_j^T\left(\frac{H_{1S_j}}{H_0}\right)^{\gamma}\right\}. \tag{14c}$$

When utility is CRRA, the term in curly braces, *ETG*, is an exact measure of the Taylor Series approximation $\mu_B\epsilon$, where the term ϵ is defined below in Equation (15).

Both $\left(\frac{H_{Ti}}{H_0}\right)$ and $\left(\frac{H_{1S_j}}{H_0}\right)$ are ratios between 0 and 1, and are magnified as γ becomes smaller, as also is $\left[\frac{H_0}{\gamma}\right]$. In CRRA, $\gamma = \omega_H$, so treatment gains are larger, the faster utility of health declines as H increases, that is, the lower the value of ω_H. This matches intuition well.

If $\gamma = 1$, which asymptotically equates to linear utility in health in CRRA, the health gain is simply $\left\{\sum \pi_i^T H_i^T - \sum \pi_j^T H_{1S_j}\right\}$, the difference in expected health levels between the treated and untreated states. Once again, this shows that CEA is a restricted version of GRACE.

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

For this specific GRACE parameter, the Taylor Series itself provides good intuition about how GRACE works. In the Taylor Series, the mean gain in HRQoL is μ_B and it is adjusted by the risk term ϵ, so the "certainty equivalent" health gain is approximately equal to $\mu_B\epsilon$. Lakdawalla and Phelps (2020) prove that

$$\epsilon = 1 + [\frac{\mu_H}{\mu_B}][-\frac{1}{2}r_H^*\Delta[\sigma_H^2][\frac{1}{\mu_H^2}] + \frac{1}{6}r_H^*\pi_H^*\Delta[\gamma_1\sigma_H^3][\frac{1}{\mu_H^3}] - \frac{1}{24}r_H^*\pi_H^*\tau^*\Delta[\gamma_2\sigma_H^4][\frac{1}{\mu_H^4}] + \ldots. \quad (15)$$

Since the risk adjustment terms are all multiplied by $\left[\frac{\mu_H}{\mu_B}\right]$, for any level of untreated illness, μ_H, they become magnified as the mean gain in treatment, μ_B, shrinks. For relatively large average treatment gains, $\epsilon \approx 1$ in many cases, but for relatively small treatment gains, ϵ can become quite important in assessing overall treatment value.

This illuminates the consequences of uncertain treatment outcomes in terms familiar to the world of finance, but converted into risk about health outcomes rather than risk about income or wealth. If utility is linear in H, so that all risk terms equal zero, then $\epsilon = 1$. Similarly, if the distributions of health outcomes are identical except for a shift in the mean, so that $\Delta\sigma_H^2$ and all higher terms show differences in the distributions of outcomes equal 0, then similarly, $\epsilon = 1$.

The term $\frac{1}{2}r_H^*\Delta\left[\sigma_H^2\right]$ is exactly parallel to the standard "risk premium" in mean-variance tradeoffs in the world of finance. It incorporates the changes in variance between treated and untreated conditions, $\Delta\sigma_H^2$, weighted by the relative risk aversion term $\frac{1}{2}r_H^*$. This term can be labeled as the "value of insurance" or the value of health risk-reduction; it is conceptually similar to the value of financial insurance that reduces risks to financial assets or income streams. If treatments reduce outcome uncertainty, $\Delta\sigma_H^2 < 0$ and value rises.

In parallel, if the distribution of treated patient outcomes has an unusually high proportion of very positive outcomes, this positive skewness in treatment outcomes also increases treated patients' expected utility. This is captured in the term $\frac{1}{6}r_H^*\pi_H^*\Delta\left[\gamma_1\sigma_H^3\right]$, which measures the changes in positive skewness of the outcome distribution, valued by the product $\frac{1}{6}r_H^*\pi_H^*$[4]. This formulation shows that positive skewness is an economic "good" for any given degree of variance. This term has been described as measuring "the value of hope," a phenomenon that has been observed in actual cancer patient preferences (Lakdawalla et al., 2012).

The next Taylor Series term encompasses changes in kurtosis of the outcome distributions. Generally, kurtosis magnifies the effect of variance in the expected utility measure.

3.1.3 TVMI in CRRA

We can now summarize TVMI as a function of d^*, t^*, ℓ^* and the CRRA power parameter γ, which combine, along with the distributions of health outcomes, to give the GRACE parameters $\rho, R,$ and D. Starting with Equation (9a) and recalling that $\mu_B\epsilon$ approximates ETG, we can write

$$TVMI_{\text{GRACE}} = KD\phi\left[\rho H_0\mu_p + ETG\omega_H R\right]. \quad (16)$$

All parameters in Equation (16) are exactly specified in CRRA, as developed within this section in Equations (11–14). The supplemental materials to this manuscript further explore

[4] The term π_H^* determines how rapidly r_H^* changes as H changes.

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

how alternative types of therapies and health states would affect GRACE valuations under CRRA utility.

3.2. HARA utility

HARA utility generalizes CRRA utility functions, which are a special case of HARA. With HARA utility

$$W(H) = [\frac{1-\gamma}{\gamma}][\frac{aH}{1-\gamma}+b]^{\gamma}. \tag{17a}$$

Since $0 \leq H \leq 1$, we have no need to scale the values of H, and therefore we can set $a = (1-\gamma)$ so:

$$W(H) = \left[\frac{1-\gamma}{\gamma}\right](H+b)^{\gamma} \tag{17b}$$

and

$$W'(H) = (1-\gamma)(H+b)^{\gamma-1}. \tag{17c}$$

Further, it is easy to prove that in HARA

$$r_H^* = (1-\gamma)\left[\frac{H}{H+b}\right] \tag{17d}$$

and

$$\omega_H = \gamma\left[\frac{H}{(H+b)}\right]. \tag{17e}$$

When $b > 0$, utility has increasing relative risk aversion (IRRA) and when $b < 0$, utility has decreasing relative risk aversion (DRRA). Obviously, when $b = 0$, then utility has CRRA.

With CRRA utility, $r_H^* + \omega_H = 1$. Introducing nonzero values of b allows this total to differ from 1.0. As we demonstrate below, this introduces a wider range of possible GRACE valuations for any value of γ, which will in turn affect all GRACE parameters. HARA generalizes CRRA by adding one extra parameter, b, increasing the formula's complexity but at the same time usefully increasing generalizability.

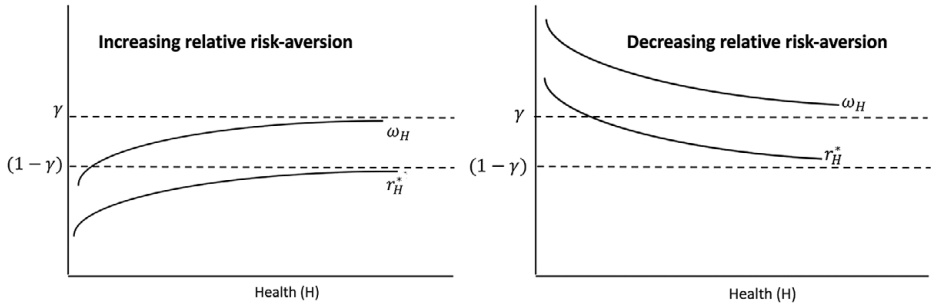

Figure 4. *Relative risk-aversion and the health elasticity of utility rise with health for IRRA utility but fall with health for DRRA utility.*

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

Figure 4(a,b) demonstrates the effects of positive and negative values of b. The intuition of these effects is reasonably obvious. When $b > 0$, as health grows, the ratios $\left[\frac{H}{H+b}\right]$ in the expressions for r_H^* and ω_H shrink for any given value of γ, but the effect diminishes as H grows. For example, if $b = 0.05$, then as H approaches 1.0, the ratio $\left[\frac{H}{H+b}\right]$ approaches 1.0 and the values of r_H^* and ω_H increase from smaller values to those near comparable CRRA values.

When $b > 0$, utility is IRRA. When the utility is IRRA, $r_H^* + \omega_H < 1$, as in Figure 4(a). As we show below, compared with CRRA utility, this generally reduces $TVMI_{GRACE}$ for any given γ and severity of acute illness or disability.

The reverse is true if $b < 0$, as in Figure 4(b). As b grows in absolute value, the ratio $\left[\frac{H}{H+b}\right]$ grows, so r_H^* and ω_H become larger, but as H grows toward 1.0 from lower values, r^* and ω_H approach the CRRA values from above.

When utility is DRRA, $r_H^* + \omega_H > 1$. As we demonstrate shortly, this combination of parameters, when compared with CRRA utility, generally makes $TVMI_{GRACE}$ larger for any given level of γ and degree of illness or disability.

The expression for R is slightly more complex, but easy to follow as an extension of Equation (11a). The levels of health in the numerator and denominator are replaced with $H + b$, so $R = \left[\frac{(H_0+b)}{H_0(1-\ell^*)+b}\right]^{1-\gamma}$. When $b > 0$, the ratio in square brackets becomes smaller as b rises, that is, as the extent of IRRA increases. This occurs because b is a larger proportion of the denominator than of the numerator.

Similarly, since $\omega_H = \gamma\left[\frac{H}{(H+b)}\right]$, when $b > 0$, the larger the value of b, the smaller is ω_H for any given value of γ. Therefore, when $b > 0$, compared with CRRA utility, two of the key parameters in K_{GRACE}, ω_H and R, shrink as b grows. Thus, before considering the disability multiplier D, the value of K_{GRACE} falls as the degree of IRRA increases.

The same is true symmetrically when $b < 0$, that is, when the utility is DRRA. Then both R and ω_H rises as the degree of DRRA increases, before considering the disability multiplier D, the value of K_{GRACE} rises as the degree of DRRA increases.

We now turn to the effects of IRRA and DRRA on the disability multiplier D. In CRRA, $D = \left[\frac{H_0}{H_0(1-d^*)}\right]^{\gamma}$, which is similar to R except that the exponent changes from $(1-\gamma)$ to γ and ℓ^* is replaced by d^*, so $D = \left[\frac{H_0+b}{H_0(1-d^*)+b}\right]^{\gamma}$. Therefore, the sign and magnitude of b have similar effects on D as they do on R. Increasing IRRA dampens the D multiplier and DRRA increases it, relative to CRRA.

Since $\rho = \sum_n \pi_n^T \left(\frac{H_n^T}{H_0}\right)^{\gamma}$ when utility is CRRA, H_n^T and H_0 are replaced by $\left(H_n^T + b\right)$ and $(H_0 + b)$, so $\rho = \sum_n \pi_n^T \left(\frac{H_n^T+b}{H_0+b}\right)^{\gamma}$. For $b > 0$, when utility is IRRA, $\left(H_n^T + b\right) < (H_0 + b)$, so ρ increases as b grows. The reverse holds and ρ diminishes as < 0, when utility is DRRA.

3.3. Exponential and EP utility

Many previous health economics studies (e.g., Zeckhauser, 1970; Feldstein, 1973; Feldstein & Friedman, 1977; Keeler et al., 1988; Manning & Marquis, 1996; Marquis & Holmer, 1996; Garber & Phelps, 1997) assumed CARA. We first discuss this simple form and then a generalization of it, EP utility (Saha, 1993).

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

3.3.1. Exponential utility

In exponential utility, which has CARA

$$U(x) = 1 - e^{-\beta x}, \tag{18a}$$

$$U'(x) = \beta x e^{-\beta x}, \tag{18b}$$

$$U''(x) = -\beta^2 x e^{-\beta x}, \tag{18c}$$

$$r = -\frac{U''(x)}{U(x)} = \beta, \tag{18d}$$

$$r^* = rx = \beta x. \tag{18e}$$

Although CARA has seen widespread use for its analytic simplicity, it has been widely rejected as appropriate in financial economics. In our setting, using CARA utility would substantially shrink $TVMI_{GRACE}$ compared with using CRRA, since CARA has strong IRRA, and the discussion of HARA with IRRA, that is, when $b > 0$, shows that all component parts of K_{GRACE} decline as IRRA becomes stronger.[5]

3.3.2. EP utility

EP utility replaces $e^{-\beta x}$ with $e^{-\beta x^\gamma}$, so that, in some sense, it is a mixture of CARA and CRRA utility. In EP utility,

$$U(x) = 1 - e^{-\beta x^\gamma}, \tag{19a}$$

$$U'^{(x)} = \gamma\beta x^{\gamma-1} e^{-\beta x^\gamma}, \tag{19b}$$

$$U''(x) = -\gamma\beta x^{\gamma-1} e^{\gamma x^{\gamma-1}} \left[\gamma\beta x^{\gamma-1}(1-\gamma)x^{-1}\right), \tag{19c}$$

$$r = -\frac{U''(x)}{U'(x)} = \frac{[\gamma\beta x^\gamma + (1-\gamma)]}{x}, \tag{19d}$$

$$r^* = rx = \gamma\beta x^\gamma + (1-\gamma). \tag{19e}$$

When $\gamma = 1$, asymptotically, relative risk aversion collapses to exponential utility, where $r^* = \beta x$. As $\gamma \to 0$ asymptotically, $r^* \to 1$, the relative risk aversion when $U(x) = \ln(x)$[6].

Unlike the CRRA and HARA functions, no simple measures exist for the key GRACE parameters such as *R, D,* and ρ.Nevertheless, it may be a very useful function to employ to characterize GRACE value measures. Just as HARA generalizes CRRA by adding one extra

[5] This would best be demonstrated in Equation (15), but the results generalize to all of our Taylor Series expansions. In CRRA, $\pi^* = r^* + 1, \tau^* = r^* + 2, \ldots$ whereas in CARA, $\pi^* = r^* = \tau^* \ldots$ Therefore, CARA necessarily shrinks all of the higher-order Taylor Series terms compared with CRRA.

[6] We use asymptotic language here because $U(x) = 1 - e^{-\beta x^\gamma}$ is invariant to x when $\gamma = 0$. In practice, EP utility is "almost CRRA" once $\gamma < 0.5$ (Phelps, 2019).

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

parameter, b, EP generalizes exponential utility by adding one parameter, γ, beyond the simpler CARA model.

4. Multiperiod model

To round out the toolkit for exact estimation of GRACE, we extend Equation (9c) to multiple periods. Define H_{Tj} as stochastic period j health in the treated sick state, H_{sj} as period j health in the untreated sick state, B_j as the period j HRQoL benefit of the treatment, and $EW_j(T) \equiv E[W(H_{Sj}+B_j) - W(H_{sj})]$. Next, define p_j^T and p_j^U as the cumulative probabilities of surviving from period zero to period j in the treated and untreated states, respectively, and $\mu_{pj} \equiv p_j^T - p_j^U$.

Finally, define β as the single-period discount factor. The period j change in utility due to the treatment can be defined as $p_j^T EW\left(H_{sj}+B_j\right) - p_j^U EW\left(H_{sj}\right)$, and the WTP for this utility change according to Equation (9a) is $\frac{K\phi H_0}{W(H_0(1-d^*))}\left\{p_j^T EW\left(H_{sj}+B_j\right) - p_j^U EW\left(H_{sj}\right)\right\}$. Therefore, the long-run period zero TVMI is

$$TVMI = \frac{\phi K H_0}{W(H_0(1-d^*))}\sum_{j=0}^{\infty}\beta^j U(C)\left\{p_j^T EW\left(H_{sj}+B_j\right) - p_j^U EW\left(H_{sj}\right)\right\}. \tag{20a}$$

Notice that the summands mechanically represent the change in expected utility between the treated and untreated states. Some algebraic manipulation transforms (20a)

$$TVMI = \frac{\phi K H_0}{W(H_0(1-d^*))}\sum_{j=0}^{\infty}\beta^j\left\{\mu_{pj}\left[EW\left(H_{sj}+B_j\right)\right] + p_j^U\left[EW_j(T)\right]\right\}. \tag{20b}$$

To put this into the context of previous GRACE parameters, we can also write Equation (20b) as

$$TVMI = \phi K D\omega_H R\sum_{j=0}^{\infty}\beta^j\left\{\frac{\mu_{pj}\left[EW\left(H_{sj}+B_j\right)\right] + p_j^U\left[EW_j(T)\right]}{W'\left(H_{sj}\right)}\right\}, \tag{20c}$$

where, of course, $KD\omega_H R = K_{\text{GRACE}}$, the WTP per unit of health gain in the full GRACE model.

Incorporating costs is straightforward. Define $Cost_j^T$ and $Cost_j^U$ as period j costs in the treated and untreated states, respectively. The expected (or per capita) incremental cost of the technology at period zero is

$$\phi\sum_{j=0}^{\infty}\beta^j\left\{p_j^T Cost_j^T - p_n^U Cost_j^U\right\}. \tag{21a}$$

Define $\Delta Cost_j \equiv Cost_j^T - Cost_j^U$. With some algebraic manipulation, we can rewrite (21a) as

$$\sum_{j=0}^{\infty}\beta^j\left\{\mu_{pj}Cost_j^T + p_j^U\Delta Cost_j\right\}. \tag{21b}$$

Combining Equations (21a) and (21b) produces the multiperiod expected net monetary benefit

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

$$NMB_{GRACE} = \phi \sum_{j=0}^{\infty} \beta^j \left\{ \mu_{pj} \left[K \frac{\left[EW\left(H_{sj} + B_j\right)\right]}{W(H_0(1-d^*))} H_0 - Cost_j^T \right] \right. \tag{22a}$$
$$\left. + p_j^U \left[K \frac{\left[EW_j(T)\right]}{W(H_0(1-d^*))} H_0 - \Delta Cost_j \right] \right\}.$$

Finally, to show the conditions under which $NMB_{GRACE} > 0$, we identify the incremental GRACE ratio. The technology should be adopted if and only if

$$K \frac{H_0}{W(H_0(1-d^*))} > \frac{\sum_{j=0}^{\infty} \left[\mu_{pj} Cost_j^T + p_j^U \Delta Cost_j \right]}{\sum_{j=0}^{\infty} \beta^j \left\{ \left[\mu_{pj} \left[EW\left(H_{sj} + B_j\right)\right] + p_j^U \left[EW\left(H_{sj} + B_j\right) - EW\left(H_{sj}\right)\right] \right] \right\}}. \tag{22b}$$

Multiplying both sides by the marginal utility of health in an arbitrary period i produces a more easily interpretable expression

$$K \frac{W'(E(H_{si}))}{W(H_0(1-d^*))} H_0 > \frac{\sum_{j=0}^{\infty} \left[\mu_{pj} Cost_j^T + p_j^U \Delta Cost_j \right]}{\sum_{j=0}^{\infty} \beta^j \left\{ \frac{W(H_0(1-d^*))}{W'(E(H_{si}))} \left[\mu_{pj} \frac{\left[EW\left(H_{sj} + B_j\right)\right]}{W(H_0(1-d^*))} + p_j^U \frac{\left[EW\left(H_{sj} + B_j\right) - EW\left(H_{sj}\right)\right]}{W(H_0(1-d^*))} \right] \right\}}. \tag{22c}$$

We note that if we multiply and divide $K \frac{W'(E(H_{si}))}{W(H_0(1-d^*))} H_0$ by $W(H_0)W'(H_0)$ in Equation (22c), then it equates to $K\omega_H R^i D$ where R^i is the severity adjustment in period i and ω_H is evaluated at $H = H_0$. Thus, the left-hand side of Equation (22c) simply represents the GRACE-adjusted to WTP, focusing on the index period i to compute the severity-adjustment R^i.

In words, this simply states that the WTP for a health gain exceeds the ratio of incremental costs to the present value of incremental benefits, all defined in terms of an exact utility function. Substitution of the CRRA utility function into these equations would provide exact measurements, a task left to the reader.

5. Eliminating discrimination according to disability

We now turn to a political question concerning the implementation of GRACE in the USA. The passage of the Affordable Care Act (ACA) in the U.S. outlawed the use of cost-effectiveness methods that discriminate against the disabled by devaluing life-years gained by people with disability. Traditional cost-effectiveness always discriminates in the sense that a marginal increase in LE is always worth less to disabled consumers, while a marginal increase in HRQoL is worth the same to the disabled and nondisabled. GRACE provides a path forward for value assessment that adheres to current US law and promotes equity for people with disability.

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

Recalling that $\mu_B \epsilon$ approximates $ETG \equiv \frac{E(W(T))}{W'(\mu_H)}$, Equation (9a) can be rewritten as

$$TVMI_{GRACE} = K\phi D\rho\mu_p H_0 + K\phi D\omega_H R\frac{E(W(T))}{W'(\mu_H)}. \tag{23}$$

The first term on the right-hand side represents the value of longevity gains, and the second is the value of HRQoL gains. Inspecting the second term, when consumers are risk averse, the marginal value of HRQoL gains is always higher for persons with disability than for those without disability. In this case, D and $W'(\mu_H)$ are both higher for the disabled, and for a technology producing a marginal increase in HRQoL, $E(W(T))$ is higher too. The other parameters in the first term do not change with disability.

The effect of disability on the first term, the value of longevity gains, is more complex. Observe that $D \equiv \frac{W(H_0)}{W(H_D)}$ is always higher for the disabled, while $\rho \equiv \frac{E(W(H_T))}{W(H_0)}$ may be lower, if the treated QoL state is lower for disabled consumers.

One case of note is that of a disability-reducing treatment. Here, ρD can exceed 1, which means extending LE has greater value for disabled people than for nondisabled persons.This situation is most easily presented using Equation (13b), with CRRA utility. Then, $\rho H_0 = H_0 \sum_n \pi_n^T \left(\frac{H_n^T}{H_0}\right)^\gamma$. Noting that each $H_n^T = H_0\left(1 - t_n^*\right)$, this becomes $\rho = \sum_n \pi_n^T \left(1 - t_n^*\right)^\gamma$. Then, $\rho D = \sum_n \pi_n^T \left[\frac{1-t_n^*}{1-d^*}\right]^\gamma$. In the simplest case, when every $t_n^* < d^*$, it follows that $\rho D > 1$. To be sure, there are cases – for example, involving negative skewness on the treatment effects – where reductions in disability might coincide with $\rho D < 1$. But for many real-life treatments, reductions in average disability severity lead to $\rho D > 1$. For instance, this condition almost certainly obtains for assistive and adaptive devices for disabled people such as wheelchairs, hearing aids, vision correction, and "public" interventions such as ramps, self-opening doors, and elevators.

More generally, Lakdawalla and Phelps (2021, 2022) present plausible conditions under which longevity gains are strictly more valuable to the disabled. However, regulators may demand an ironclad guarantee of nondiscrimination. To solve this problem, analysts may proceed just like they do with income inequity. More specifically, just as analysts routinely calculate WTP for health improvement under the assumption that income is uniform, one may analogously calculate WTP under the assumption that the initial health level is uniform too.

Two approaches may be pursued. The first guarantees "ex ante" equity, by calculating ρD under the assumption that all consumers start out with perfect ex ante health, H_0. In this case, $\rho D = \frac{E(W(H_T))}{W(H_0)}$, and H_T would be the same for the disabled and nondisabled. However, it would still be true that longevity gains would be worth less to those suffering more severe illness.

To address the latter point, one may adopt a more aggressive solution that also guarantees "ex post" equity, by assuming that post-treatment health levels are uniform across all patients. This would amount to using $\rho D = 1$.

6. Parameter estimation

We know of two general approaches to estimate the necessary parameters: discrete choice experiments (DCEs) and "happiness economics" regression approaches. We discuss these in turn.

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

6.1. Discrete choice experiments

We begin with the simplest case, that of CRRA utility. Holt and Laury (2002) demonstrate the key methodology, as shown in Table 1.

In this approach, people are presented with pairs of gambles that alter the risk/reward tradeoffs. This can be done either by altering the payoffs or the probabilities of the gambles; Holt and Laury modified the probabilities, while Noussair et al. (2014) modified the gambles with constant probabilities.

In Table 1, Option A, the "safe" choice, is less-risky than Option B. Moving down the rows, as the probabilities shift, the expected payoff steadily increases, so that choosing the riskier gamble, Option B, becomes increasingly attractive. At some point, respondents switch from making "safe" choices to the riskier choices. Table 2 uses CRRA to determine the expected utility of each gamble, and the switching point identifies the CRRA power parameter γ. Once this is known, and CRRA is assumed, the exact utility function expressions for TVMI are fully defined. For example, in Table 2, if the switch occurs at the fifth gamble, then relative risk aversion is somewhere between 0.15 and 0.41.

Only 8 per cent of the subjects exhibited risk-loving behavior (the top three rows), and two-thirds were clearly risk averse. The median number of safe choices was 5, indicating relative risk aversion in the range of $.15 < r^* < .41$, with a midpoint of $r^* \approx .28$.

Holt and Laury (2002) also estimated an EP model using their data, where, at mean values of the data, the estimated value of $r^* \approx .28$ and climbs slowly with the size of the gamble.

The same MLE approach can obviously be used for any specified utility function, although the complexity will increase with the number of parameters necessary to define utility. Just as with EP utility, HARA utility requires two parameters, the power parameter γ and the risk aversion adjustment b.

Table 1. *Discrete choice experiment structure using financial gambles.*

Option A		Option B		Expected payoff difference
\$2.00 with $p = 0.1$	\$1.60 with $p = 0.9$	\$3.85 with $p = 0.1$	\$0.10 with $p = 0.9$	\$1.17
\$2.00 with $p = 0.2$	\$1.60 with $p = 0.8$	\$3.85 with $p = 0.2$	\$1.60 with $p = 0.8$	\$0.83
\$2.00 with $p = 0.3$	\$1.60 with $p = 0.7$	\$3.85 with $p = 0.3$	\$1.60 with $p = 0.7$	\$0.50
\$2.00 with $p = 0.4$	\$1.60 with $p = 0.6$	\$3.85 with $p = 0.4$	\$1.60 with $p = 0.6$	\$0.16
\$2.00 with $p = 0.5$	\$1.60 with $p = 0.5$	\$3.85 with $p = 0.5$	\$1.60 with $p = 0.5$	−\$0.18
\$2.00 with $p = 0.6$	\$1.60 with $p = 0.4$	\$3.85 with $p = 0.6$	\$1.60 with $p = 0.4$	−\$0.51
\$2.00 with $p = 0.7$	\$1.60 with $p = 0.3$	\$3.85 with $p = 0.7$	\$1.60 with $p = 0.3$	−\$0.85
\$2.00 with $p = 0.8$	\$1.60 with $p = 0.2$	\$3.85 with $p = 0.8$	\$1.60 with $p = 0.2$	−\$1.18
\$2.00 with $p = 0.9$	\$1.60 with $p = 0.1$	\$3.85 with $p = 0.9$	\$1.60 with $p = 0.1$	−\$1.52
\$2.00 with $p = 1$	\$1.60 with $p = 0$	\$3.85 with $p = 1$	\$1.60 with $p = 0$	−\$1.85

Notes: From Holt and Laury (2002, Table 1). Option A has very low risk. Option B has higher risk. The final column shows the expected difference between choosing Option A over Option B.

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

Table 2. *Summary of risk aversion for CRRA utility.*

Number of safe choices	Range of relative risk aversion (r^*) using CRRA utility function	Descriptive phrase
0–1	Lower than −0.95	Highly risk loving
2	−0.95 to −0.49	Very risk loving
3	−0.49 to −0.15	Mildly risk loving
4	−0.15 to 0.15	Risk neutral
5	0.15 to 0.41	Mildly risk averse
6	0.41 to 0.68	Risk averse
7	0.68 to 0.97	Very risk averse
8	0.97 to 1.37	Highly risk averse
9 to 10	>1.37	Stay in bed!

Note: From Holt and Laury (2002, Table 3).

6.2. Happiness economics models

While DCE experiments focus on gambles involving health states, Happiness Economics models offer the ability to simultaneously estimate not only the parameters for the "health" components of GRACE, but also the elasticity of utility with respect to consumption, ω_C in the same data set as is used to estimate the functional form and parameters of $W(H)$.

Easterlin (2003) pioneered the Happiness economics approach, now widely used by the United Nations using data from international Gallup polls. This approach requires (in our case) three data elements from each respondent: (i) Their level of happiness on some fixed scale, for example, 0–10 or 0–100; (ii) their income, and (iii) their level of health on a fixed interval, for example, 0–10 or 0–100. These can be gathered using direct questions, visual analog scales ("thermometer" scales) or other methods. This approach uses reported happiness as a proxy for the economists' concept of "utility." With such data, one can estimate a generic translog utility function (Christiansen et al., 1973) of the form

$$\ln(Happy_i) = \beta_1 \ln(H_i) + \frac{1}{2}\beta_2(\ln(H_i))^2 + \beta_3 \ln(C_i) + \frac{1}{2}\beta_4(\ln(C_i))^2 + \epsilon_i \ldots. \qquad (24)$$

Of course, analysts can include other covariates that might affect happiness such as age, sex, ethnicity, geographic region, and others. These would reduce residual variance, and would avoid omitted variable bias if the omitted variables were correlated with H or C.

From such data, one can readily infer the key GRACE parameters such as ω_H, r_H^* (and how these change with levels of H), and ω_C and how it might change with C. Phelps and Lakdawalla (2023, Chapter 8) provide details for this approach.

Assuming a specific utility function offers a different approach, however, namely to fit the data from the "happiness" survey to a specific functional form. We use here the example of HARA utility, although the approach generalizes to any specific form of the utility function. Here again, the assumption of separability in utility simplifies the discussion.

Begin with the standard utility function in Equation (3), but with the specific assumption that both $U(C)$ and $W(H)$ have HARA form. In what follow, we will ignore the possible inclusion of covariates, which can generically be added as X_j values, each with their own

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

appropriate functional form and parameterization. We add a generic error term, the details of which would emerge in econometric analysis. For individual j,

$$\text{Happiness} = U\left(C_j\right)W\left(H_j\right)\epsilon_j. \tag{25a}$$

Applying CRRA utility from Equation (17b) gives

$$Happiness_j = \left[\frac{1-\gamma}{\gamma}\right]\left[C_j+b\right]^{\gamma}\left[\frac{1-\delta}{\delta}\right]\left[H_j+\beta\right]^{\delta}\epsilon_{\mathrm{j}}. \tag{25b}$$

In logarithmic form,

$$\ln\left(Happiness_j\right) = \ln\left(\frac{1-\gamma}{\gamma}\right) + \ln\left(\frac{1-\delta}{\delta}\right) + \gamma\ln\left[C_j+b\right] + \delta\ln\left[H_j+\beta\right] + \ln\left(\epsilon_j\right). \tag{25c}$$

From this, the log-likelihood function is readily composed, and maximized over the four parameters $\gamma, b, \delta,$ and β, plus those associated with any other covariates included in the model. One could readily test the hypotheses that the functions $U(C)$ and $W(H)$ were CRRA by testing whether (respectively) b and β equal zero. Econometric analysis would determine the best possible transformation for ϵ_j, possibly using Box and Cox (1964) transformations to normalize the distribution of the residuals. We leave these details to others.

One could obviously insert the EP function into Equation (25b) as an alternative to the HARA function, which would have a more complicated appearance, but would still only require maximization over two parameters each for $U(C)$ and $W(H)$.

Adding covariates such as age, sex, ethnicity, and others could improve precision by reducing the residual variance. One could compare across structural models by comparing log-likelihood ratios from equations using HARA and EP. Unfortunately, HARA and EP are not nested, so they cannot be compared in a single maximum likelihood estimation.

7. Conclusion

Previous presentations of GRACE have relied on Taylor series expansion methods to develop estimates of the necessary GRACE parameters. Herein, we have developed the GRACE model while assuming a specific utility function to replace the Taylor Series coefficient estimates. The simple CRRA model seems to provide the most intuitive understanding of how the GRACE model adjusts WTP for three key GRACE parameters: the rate at which utility changes with health, ω_{H}, the disability adjustment, D, and the illness severity adjustment, R.

Practitioners of CEA can now choose between the Taylor Series model and the exact utility function model, depending on the nature of data available to them. We emphasize here that these parameters need not be estimated for every health technology assessment (HTA) or study. Once the research community has come to some agreement about the values of the key parameters, whether those in Taylor Series expansions or for specific utility functions, then those parameters can be used in every HTA study. The normal data collected in such studies, such as in randomized controlled trials and clinical assessments of health technologies suffice to complete the models by providing information on the distributions of health outcomes for treated and untreated individuals, as summarized in Equation (16) in the simple two-period model and in Equation (20b) in the complete multiperiod model.

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

We have not dwelt herein on estimation of costs in HTA studies since the methods are the same in our model as for those in standard CEA, and indeed, the same as would take place in any BCA model of value for healthcare. Similarly, issues of discounting do not differ from standard CEA and BCA models. GRACE focuses on new ways to measure value. This article summarizes a new approach to measuring the differences in value for new medical interventions, be they treatments, diagnostic tools, preventive medical interventions such as vaccines, or public health measures such as provision of sanitary water.

This approach informs how BCA studies of medical interventions should be undertaken, since BCA and CEA share similar roots. Indeed, many CEA practitioners now routinely report their outcomes in terms of net monetary benefit (NMB) measures, which require assuming a specific value for *K,* the WTP for improved health. The "near" equivalence of CEA and BCA has been long-understood (Phelps & Mushlin, 1991) and the specific conditions under which this equivalence arises have been fully developed elsewhere (Bleichrodt & Quiggin, 1999). The GRACE model, with separable utility and additive utility over time, fulfills these conditions.

If analysts wish to conduct BCA studies for health technologies, GRACE shows that measuring aggregate health benefits in such studies does not suffice. When utility is linear in health, as CEA assumes, then adding up health benefits across individuals is legitimate, but when there are diminishing returns to health, then analyses must incorporate such measures as untreated illness severity and severity of preexisting disabilities to be valid, and should also account for uncertainty in treatment outcomes and a MRS between LE and HRQoL that depends upon baseline conditions.

Parameter Definitions

$H(a,b)$	the production function for producing health using a and b
H'_a and H'_b	the marginal products of a and b in producing H
p_a and p_b	the market prices for health inputs a and b
C	consumption, equals income minus medical care spending
H	health, measuring the health-related quality of life (HRQoL)
$U(C)$ and $U'(C)$	utility and marginal utility of C
ω_C	the elasticity of utility with respect to C
$W(H)$ and $W'(H)$	the utility and marginal utility of H
ω_H	the elasticity of utility with respect to H
$V(C, H)$	$V(C,H) = U(C)\ W(H)$, the separable combined utility function
K	the standard CEA measure of WTP for QALY gains
H_0	baseline health, on a scale of $0 \leq H \leq 1$, $H_0 = 1$, "perfect" health
ℓ^*	on a scale of $0 \leq \ell^* \leq 1$, proportional health loss from untreated illness
H_{1S}	$H_{1S} = H_0(1 - \ell^*)$, the level of untread health with an illness
μ_H	the mean of H_{1S}
d^*	on a scale of $0 \leq d^* \leq 1$, the proportional loss in health from disability
H_D,	$H_D = H_0(1 - d^*)$, the level of health with a permanent disability
R	adjustment to WTP for illness severity; $R = \frac{W'(\mu_H)}{W'(H_0)}$
D	adjustment to WTP for disability; $D = \frac{W(H_0)}{W(H_D)}$
μ_B	average gain (benefit) in health from treatment, in units of H
μ_P	average gain in LE from treatment, in units of probability of survival

https://doi.org/10.1017/bca.2023.6 Published online by Cambridge University Press

K^D_{GRACE}	WTP after GRACE adjustment; $K^D_{GRACE} = DR\omega_H K$
$EW(T)$	expected treatment gain, in units of W(H)
ETG	expected value of treatment gains, measured in units of H; $ETG = \frac{EW(T)}{W'(\mu_H)}$
ϕ	probability of illness occurring in period 1
ϵ	adjustment for treatment outcome uncertainty in Taylor Series
ρ	ρH_0 is the MRS between LE and HRQoL

Acronyms

CARA	constant absolute risk aversion
CEA	standard cost-effectiveness analysis
CRRA	constant relative risk aversion
DRRA	decreasing relative risk aversion
EP	expo-power utility, a generalization of CARA utility
HARA	hyperbolic absolute risk aversion, a generalization of CRRA
HRQoL	health-related quality of life
ICER	incremental cost-effectiveness ratio
IRRA	increasing relative risk aversion
LE	life expectancy
MRS	marginal rate of substitution
QALY	quality adjusted life-year – a standard measure of amounts of health
TVMI	total value of a medical intervention
WTP	willingness to pay, measured in terms of C per unit of H

Supplementary Materials. To view supplementary material for this article, please visit http://doi.org/10.1017/bca.2023.6.

References

Bleichrodt, Han, and John Quiggin. 1999. "Life-Cycle Preferences over Consumption and Health: When Is Cost-Effectiveness Analysis Equivalent to Cost–Benefit Analysis?" *Journal of Health Economics*, *18*: 681–708.

Box, George E. P., and David R. Cox. 1964. "An Analysis of Transformations." *Journal of the Royal Statistical Society, Series B*, *26*(2): 211–252.

Christiansen, Laurits R., Dale W. Jorgensen, and Lawrence J. Lau. 1973. "Transcendental Logarithmic Production Frontiers." *Review of Economics and Statistics*, *25*(1): 28–45.

Cordoba, Juan, and Maria Ripoll. 2017. "Risk Aversion and the Value of Life." *Review of Economic Studies*, *84*(4): 1472–1509.

Easterlin, Richard A. (2003). "Explaining happiness." *Proceedings of the National Academy of Sciences 100*(19): 11176–11183.

Feldstein, Martin S. 1973. "The Welfare Loss of Excess Health Insurance." *Journal of Political Economy*, *81*(2): 251–280.

Feldstein, Martin, and Bernard Friedman. 1977. "Tax Subsidies, the Rational Demand for Insurance and the Health Care Crisis." *Journal of Public Economics*, *7*(2): 155–178.

Garber, Alan M., and Charles E. Phelps. 1997. "The Economic Foundations of Cost-Effectiveness Analysis." *Journal of Health Economics*, *16*(1): 1–31.

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review*, *95*(5): 1644–1655.

Holt, Charles A., and Susan K. Laury. 2014. "Assessment and Estimation of Risk Preferences." *Handbook of the Economics of Risk and Uncertainty*, *1*: 135–201.

Keeler, Emmett B., Joan L. Buchanan, John E. Rolph, Janet M. Hanley, and David Reboussin. 1988. *The Demand for Episodes of Treatment in the Health Insurance Experiment*. Santa Monica, CA: RAND Corporation.

Kimball, Miles S. 1990. "Precautionary Saving in the Small and in the Large." *Econometrica*, *58*(1): 53–73.

Lakdawalla Darius, N., and Charles E. Phelps. 2020. "Technology Evaluation with Risk Aversion in Health." *Journal of Health Economics*, *72*: 102346.

Lakdawalla Darius, N., John A. Romley, Yuri Sanchez, J. Ross Maclean, John R. Penrod, and Tomas Phillipson. 2012. "How Cancer Patients Value Hope and the Implications for Cost-Effectiveness Assessments of High-Cost Cancer Therapies." *Health Affairs*, *31*: 676–682.

Lakdawalla, Darius N., and Charles E. Phelps. 2021. "Health Technology Assessment with Diminishing Returns to Health: The Generalized Risk-Adjusted Cost-Effectiveness (GRACE) Approach." *Value in Health*, *24*(2): 1244–1249.

Lakdawalla, Darius N., and Charles E. Phelps. 2022. "A Guide to Extending and Implementing Generalized Risk-Adjusted Cost-Effectiveness (GRACE)." *European Journal of Health Economics*, *23*: 433–451. https://doi.org/10.1007/s10198-021-01367-0.

Manning, Willard G., and M. Susan Marquis. 1996. "The Trade-Off Between Risk Pooling and Moral Hazard." *Journal of Health Economics*, *15*(5): 609–639.

Marquis, M. Susan, and Martin R. Holmer. 1996. "Alternative Models of Choice Under Uncertainty and the Demand for Health Insurance." *Review of Economic Studies*, *78*(3): 421–427.

Noussair, Charles N., Stefan T. Trautmann, and Gijs Van de Kuilen. 2014. "Higher Order Risk Attitudes, Demographics, and Financial Decisions." *Review of Economic Studies*, *81*: 325–355.

Phelps, Charles E. 2019. "A New Method for Determining the Optimal Willingness to Pay in Cost-Effectiveness Analysis." *Value in Health*, *22*(7): 785–791.

Phelps, Charles E., and Chris Cinatl. 2021. "Estimating Optimal Willingness to Pay Thresholds for Cost-Effectiveness Analysis: A Generalized Method." *Health Economics*, *30*(7): 1697–1702.

Phelps, Charles E., and Darius N. Lakdawalla. 2023. *Valuing Health: The Generalized Risk-Adjusted Cost Effectiveness (GRACE) Model*. New York: Oxford University Press.

Phelps, Charles E., and Alvin I. Mushlin. 1991. "On the (Near) Equivalence of Cost-Effectiveness and Cost-Benefit Analysis." *International Journal of Health Technology Assessment*, *7*(1): 12–21.

Saha, Atanu. 1993. "Expo-Power Utility: A 'Flexible' Form for Absolute and Relative Risk Aversion." *American Journal of Agricultural Economics*, *75*(4): 905–913.

Taylor, Brook. 1715. *Methodus incrementorum directa & inversa*. London: Gulielmi Innys.

Williams, Alan. 1992. "Cost-Effectiveness Analysis: Is it Ethical?" *Journal of Medical Ethics*, *18*: 7–11.

Zeckhauser, Richard. 1970. "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives." *Journal of Economic Theory*, *2*: 1–26.

Cite this article: Lakdawalla, D. N. and Phelps, C. E. 2023. "The Generalized Risk-Adjusted Cost-Effectiveness (GRACE) Model for Measuring the Value of Gains in Health: An Exact Formulation." *Journal of Benefit-Cost Analysis* 14: 44–67, doi:10.1017/bca.2023.6