



## ORIGINAL ARTICLE

# Expected utility versus expected regret theory versions of decision curve analysis do generate different results when treatment effects are taken into account

Iztok Hozo PhD<sup>1</sup> | Athanasios Tsalatsanis PhD<sup>2,4</sup> | Benjamin Djulbegovic MD, PhD<sup>3,5,6,7</sup>

<sup>1</sup> Professor and Chair, Department of Mathematics, Indiana University Northwest, Gary, Indiana, USA

<sup>2</sup> Associate Professor, USF Health Program for Comparative Effectiveness Research, Tampa, Florida, USA

<sup>3</sup> Distinguished Professor and Associated Dean, USF Health Program for Comparative Effectiveness Research, Tampa, Florida, USA

<sup>4</sup> Associate Professor, Division for Evidence Based Medicine, Department of Internal Medicine, University of South Florida, Tampa, Florida, USA

<sup>5</sup> Distinguished Professor and Associate Dean, Division for Evidence Based Medicine, Department of Internal Medicine, University of South Florida, Tampa, Florida, USA

<sup>6</sup> Departments of Hematology and Health Outcomes Behavior, H. Lee Moffitt Cancer and Research Institute, Tampa, Florida, USA

<sup>7</sup> Tampa General Hospital, Tampa, Florida, USA

**Correspondence**

Benjamin Djulbegovic, 12901 Bruce B. Downs Blvd, MDC02, Tampa, FL 33612, USA.  
Email: bdjulbeg@health.usf.edu

**Funding information**

DOD, Grant/Award Number: W81 XWH 09-2-0175 (PI: Djulbegovic)

**Abstract**

**Rationale, aims, and objectives** Decision curve analysis (DCA) is a widely used method for evaluating diagnostic tests and predictive models. It was developed based on expected utility theory (EUT) and has been reformulated using expected regret theory (ERG). Under certain circumstances, these 2 formulations yield different results. Here we describe these situations and explain the variation.

**Methods** We compare the derivations of the EUT- and ERG-based formulations of DCA for a typical medical decision problem: “treat none,” “treat all,” or “use model” to guide treatment. We illustrate the differences between the 2 formulations when applied to the following clinical question: at which probability of death we should refer a terminally ill patient to hospice?

**Results** Both DCA formulations yielded identical but mirrored results when treatment effects are ignored; they generated significantly different results otherwise. Treatment effect has a significant effect on the results derived by EUT DCA and less so on ERG DCA. The elicitation of specific values for disutilities affected the results even more significantly in the context of EUT DCA, whereas no such elicitation was required within the ERG framework.

**Conclusion** EUT and ERG DCA generate different results when treatment effects are taken into account. The magnitude of the difference depends on the effect of treatment and the disutilities associated with disease and treatment effects. This is important to realize as the current practice guidelines are uniformly based on EUT; the same recommendations can significantly differ if they are derived based on ERG framework.

**KEYWORDS**

evaluation, medical research, practical reasoning

## 1 | INTRODUCTION

Arguably, the threshold model represents one of the most important advances in clinical decision making.<sup>1,2</sup> According to the threshold model when faced with uncertainty about whether to treat, order a test or apply a predictive model, or simply observe the patient, there exists some probability of disease or disease outcome (threshold), at which a physician is indifferent between administering versus not administering

treatment, or acting according to test or predictive model.<sup>1,2</sup> The threshold model reflects one of the fundamental principles of rational decision making: it is rational for a doctor to act (ie, order a diagnostic test, prescribe treatment) and for the patient to accept the proposed health intervention when one believes that benefits (gains) (B) of such action will outweigh its harms (losses) (H), ie, exceed threshold (T).<sup>3,4</sup>

Originally, the threshold model was derived within the precepts of the expected utility theory (EUT).<sup>1,2</sup> During the last 40 years, the

threshold model has been reformulated in a number of ways, both within the framework of EUT and non-EUT theories (for a review, see Djulbegovic et al<sup>4</sup>). One such extension of the threshold model is decision curve analysis (DCA).

DCA is a widely used technique for evaluation of the value of diagnostic tests or predictive models over a range of all possible thresholds.<sup>5-10</sup> The assessment of the threshold probability at which a decision maker is indifferent between failure to administer a beneficial, over committing to a potentially harmful health intervention, allows capturing patient preferences related to given management choices.<sup>1,2,4</sup> DCA incorporates the predictive model's accuracy, the consequences of a decision action, and a patient's preferences to assess the best course of action, such as making a decision according to the predictive model, treat all patients, or treat none.<sup>5-7</sup> One of the advantages of DCA is that we do not actually need to elicit the threshold from each patient, but instead model decisions about treatment over a range of thresholds without knowing details about specific utilities that determine threshold.<sup>5,6</sup>

DCA was originally formulated using EUT<sup>5,6</sup> and reformulated within the expected regret theory (ERG) framework.<sup>7,11</sup> In the EUT DCA, the best course of action is the one associated with the highest expected value, whereas in the ERG DCA, the best course of action is the decision that will lead to the least amount of regret.<sup>7</sup>

We previously showed that ERG DCA and EUT DCA lead to the same decisions if treatment effects are not taken into consideration.<sup>7</sup> Note, however, that the original DCA did not explicitly model treatment effect on patient's outcomes. In this paper, we demonstrate that when treatment effects are included in the modeling, different results are generated by EUT and ERG DCA, which has important implications for medical decision making.

## 2 | METHODS

### 2.1 | Model structure

Figure 1 displays a typical decision tree describing treatment options based on the results of a prediction model.  $q_i$  represents the model-

generated probability of the event of interest  $D$  ( $D+$  event present,  $D-$  event absent), such as disease presence or occurrence of outcome in a patient  $i$ ;  $p_i$  is the actual probability of the event  $D+$  for the patient  $i$ . The individual's risk is  $p_D = \frac{1}{N} \sum p_i$ , where  $N$  is the number of patients. RRR is the relative risk reduction expected from treatment Rx,  $U_j$  is the utility associated with each outcome  $j$ , and  $T$  is a threshold probability at which a decision maker is indifferent between the "do not treat" (NoRx) and "treat" (Rx) strategies.

We opt to model treatment effect as relative risk reduction (RRR), which is a convenient way to express a risk ratio (RR) as a proportion of risk ( $p_i$ ) reduction according to<sup>15,16</sup>

$$RRR = 1 - RR,$$

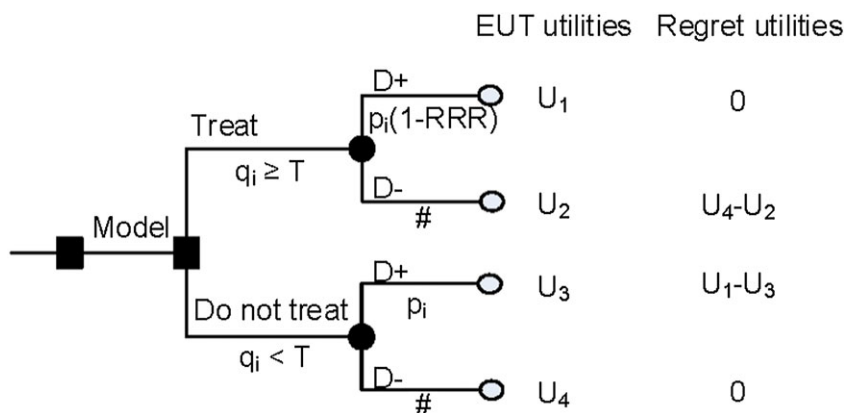
where RR is defined as the risk of event in the treatment group over the risk of event in control group.<sup>16</sup>

The main advantage of using RRR as a measure of treatment effect over the treatment absolute differences is that the former remains constant over the range of predicted risks ( $p_i$ ).<sup>15-17</sup> RRR is also easy to interpret: RRR = 1 means that the occurrence of outcome of interest is completely preventable [as  $p(1 - RRR) = 0$ ], whereas RRR = 0 means that treatment is useless as it does not affect underlying risk [ $p(1 - 0) = p$ ].

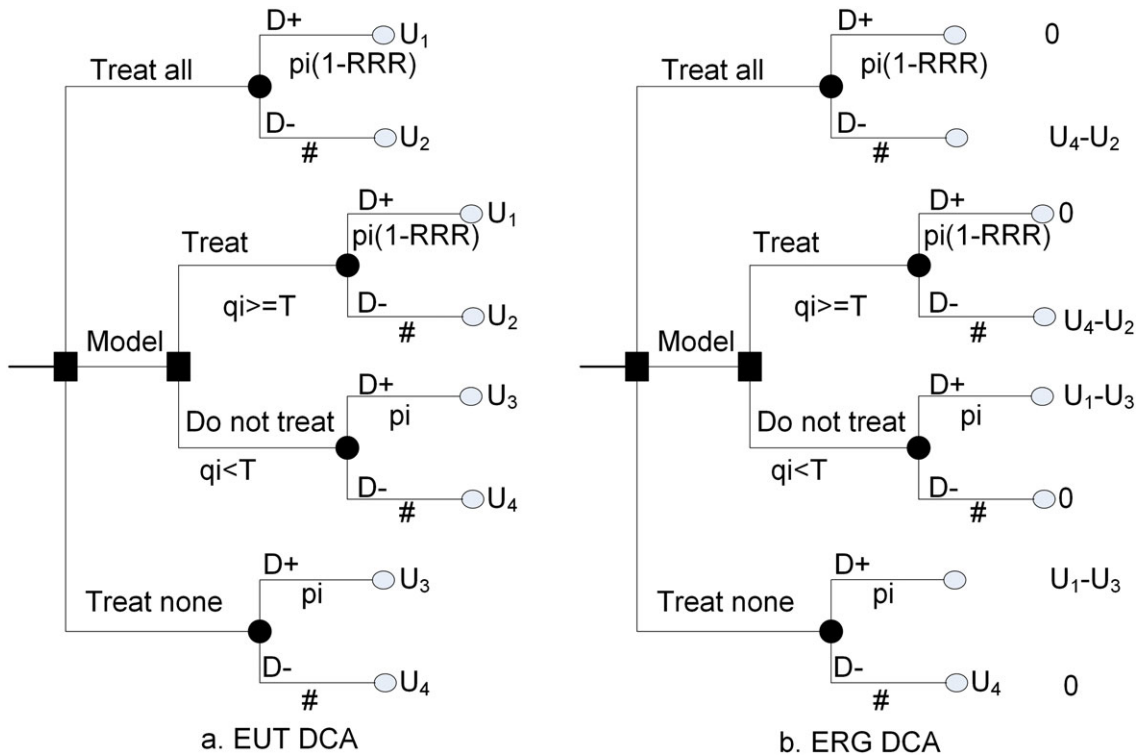
Note that we use the term "predictive model" in generic sense to predict or foresee/foretell something that is yet unknown (such as outcome occurrence in individual patients). Typically, such models convert available information (predictors) into a statement about the probability about diagnosis or prognosis.<sup>18,19</sup> The model shown in Figure 1 (and later in Figure 2) applies to both prognostic and diagnostic prediction as long as such a prediction is used to guide selection of treatment.

### 2.2 | Derivation of the EUT DCA

As an illustrative example, Figure 2a shows the decision tree of a 3-choice dilemma associated with hospice referral. In this case, a patient may decide to receive treatment targeting his underlying disease ("Treat All"), accept referral to hospice ("Treat None"), or act according to the threshold model based the patient's estimated probability of



**FIGURE 1** Decision tree depicting use of a predictive model to guide treatment choices according to expected utility-based theory ("EUT utilities") and ERG (displayed as "Regret Utilities," ie, differences in values or utility of the outcomes of the action taken and the utility of the outcomes of another action, which, in retrospect, we should have taken<sup>12-14</sup>). As explained in the text,  $q_i$  represents the model-generated probability of disease outcome for a patient  $i$ , whereas  $p_i$  is the actual probability of the event  $D$  (disease outcome) for the same patient.  $T$ , threshold probability for treatment;  $D+$ , disease is present;  $D-$ , disease is absent; RRR, relative risk reduction of treatment. Regret is computed as the difference in utilities of the action taken and the action that, in retrospect, should have been taken



**FIGURE 2** Decision tree depicting a typical 3-choice dilemma. a, Expected utility-based tree. b, expected regret-based tree. Three alternatives are shown: treat all patients, treat none, and use a predictive model/test to decide whether to treat or not.  $q_i$  represents the model-generated probability of disease outcome for a patient  $i$  whereas  $p_i$  is the actual probability of the event D+ (disease outcome) for the same patient. T, threshold probability for treatment; D-, disease is absent; RRR, relative risk reduction of treatment. Regret is computed as the difference in utilities of the action taken and the action that, in retrospect, should have been taken

death ("Model"). In Figure 2a,  $q_i$  and  $p_i$  represent the model estimated probability of death and the actual probability of death D+ for patient  $i$  respectively. RRR represents the relative risk reduction associated with treatment Rx,  $U_j$  is the utility of outcome  $j$ , and  $T$  is the threshold probability at which a decision maker is indifferent between benefits and harms of treatment.

By solving the tree, we derive the expected value of the model as

$$E[\text{Model}] = \frac{1}{N} \left( \sum_{q_i \geq T} [p_i(1-RRR)U_1 + (1-p_i)(1-RRR)U_2] + \sum_{q_i < T} [p_i U_3 + (1-p_i)U_4] \right) \tag{1}$$

As explained earlier, according to the threshold model, we treat if  $q_i \geq T$ . Therefore, the expected values of the "Treat none" and "Treat all" strategies can be computed by setting  $T = 1$  and  $T = 0$  in Equation 1, respectively. Thus,

$$E[Rx \text{ none}] = [p_D U_3 + (1-p_D)U_4],$$

$$E[Rx \text{ All}] = [p_D(1-RRR)U_1 + (1-p_D(1-RRR))U_2].$$

The optimal strategy is the one that yields the higher net benefit (NB). Using the derived expected utilities, we calculate the NB of a strategy by subtracting the expected utility of this strategy (eg, "Treat all" or "Model") from the expected utility of the "Treat none" strategy.<sup>5,6</sup> Thus,

$$NB[\text{Model}] = \frac{1}{N} \left( \sum_{q_i \geq T} [p_i(U_1-U_3) - (1-p_i)(U_4-U_2) - p_i(RRR)(U_1-U_2)] \right) \tag{2}$$

To simplify Equation 2, we define the true positive rate for a given threshold  $T$  as

$$TP_T = \frac{1}{N} \sum_{q_i \geq T} p_i,$$

and False Positive rate for the given threshold  $T$  as

$$FP_T = \frac{1}{N} \sum_{q_i \geq T} (1-p_i).$$

Thus,

$$NB[\text{Model}] = TP_T \cdot (U_1-U_3) - FP_T \cdot (U_4-U_2) - TP_T \cdot RRR \cdot (U_1-U_2).$$

By replacing  $T = 0$  in Equation 2, we derive the NB of "Treat all" ( $P_T = p_D$  and  $FP_T = 1 - p_D$ ),

$$NB[Rx \text{ All}] = p_D \cdot (U_1-U_3) - (1-p_D) \cdot (U_4-U_2) - p_D \cdot RRR \cdot (U_1-U_2).$$

To further simplify our notation, and following Pauker and Kassirer<sup>1,2</sup> and Vickers and Elkin,<sup>5</sup> we define the differences between utilities (preferences) related to the consequence of administering treatment when it would have been of benefit as  $B = U_1 - U_3$ ; similarly, the preferences related to the consequences of being unnecessarily treated are denoted as harms  $H = U_4 - U_2$ . Finally, even if appropriately

given, there is no guarantee that only ( $D+$ ) patients will receive treatment (see Figure 2); some patients with ( $D-$ ) may also receive treatment. We defined this difference ( $\Delta$ ) between utilities of administering treatment as  $\Delta = U_1 - U_2$ . Note that all threshold models to date assumed that the differences in NB and harm between these utilities are positive ( $B \geq 0$  and  $H \geq 0$ ), which is a clinically sensible assumption. Although, in principle, that is possible, we do not consider the case of negative utilities or negative  $B$  or  $H$  in our threshold model.

With these substitutions, we can rewrite the formulas above as

$$\begin{aligned} \text{NB}[\text{Model}] &= TP_T \cdot B - TP_T \cdot \text{RRR} \cdot \Delta - FP_T \cdot H \\ &= TP_T \cdot (B - \text{RRR} \cdot \Delta) - FP_T \cdot H, \end{aligned} \quad (3)$$

$$\begin{aligned} \text{NB}[\text{Rx All}] &= p_D \cdot B - p_D \cdot \text{RRR} \cdot \Delta - (1 - p_D) \cdot H \\ &= p_D \cdot (B - \text{RRR} \cdot \Delta) - (1 - p_D) \cdot H. \end{aligned} \quad (4)$$

The threshold probability, or the probability at which one is indifferent between deciding to treat versus not to treat, is computed as

$$T \cdot U_3 + (1 - T) \cdot U_4 = T \cdot (1 - \text{RRR}) \cdot U_1 + (1 - T) \cdot (1 - \text{RRR}) \cdot U_2.$$

Using the definitions for  $B$ ,  $H$ , and  $\Delta$ , above, we have

$$H = T \cdot (B + H - \text{RRR} \cdot \Delta), T_{EUT} = \frac{H}{B + H - \text{RRR} \cdot \Delta} \quad (5)$$

Note that if  $\text{RRR} = 0$ ,  $T_{EUT}$  reduces to the "classic" EUT Pauker and Kassirer threshold<sup>1,2</sup>:

$$T_c = \frac{H}{B + H}.$$

$B$ ,  $H$ , and  $\Delta$  can be further characterized in terms of disutilities, or using other popular evidence-based statistical measures.<sup>20,21</sup> When done so, the threshold model can be further formulated in a number of other ways (for a review, see Djulbegovic et al<sup>4</sup>).

Although the original and widely used DCA did not take treatment effect into consideration, Vickers et al<sup>10</sup> did attempt to integrate treatment into EUT DCA by expressing the threshold as the absolute risk reduction between 2 treatments:  $\text{ARD} = p_0 - p_t$ , where " $p_t$ " represents the probability of event for patients receiving treatment and " $p_0$ " represents the probability of event in untreated patients". In our views, this creates several problems. First,  $\text{ARD}$  does vary with baseline risk, and it is preferable to model treatment effects using relative effects such as  $\text{RRR}$  that remain constant over the range of predicted risks ( $p_i$ ).<sup>15-17</sup> Second, it may be better to express the threshold with respect to individual risk probabilities ( $p_i$ ). In principle, that is possible by expressing  $p_t = p_0 \cdot (1 - \text{RRR})$ , which would, in turn, allow reformulation of threshold via  $p_0$ . However, most importantly, the method described by Vickers et al<sup>10</sup>, assume that  $U_1 - U_3 < 0$ . Although technically this is correct, clinically such a situation constitute an extremely rare case, particularly in the area of cancer treatment discussed by the authors. Regardless if one wants to apply this EUT DCA model, its use per se is immaterial to the main objective of our paper, which is to contrast findings using EUT DCA with ERG DCA.

In further exposition, we do not consider all possible ways how the threshold equation can be expressed but focus on derivation of DCA, which is mainly done by scaling the original threshold formulas. As

explained, our main intent here is to demonstrate differences between EUT- and ERG-derived DCA. Using the relationship expressed in Equation 4, we can derive scaled NBs as

Scaled by  $B - \text{RRR} \cdot \Delta$

$$\text{NB}_{B-\text{RRR} \cdot \Delta}[\text{Model}] = TP_T - FP_T \cdot \frac{H}{B - \text{RRR} \cdot \Delta} = TP_T - FP_T \cdot \frac{T_{EUT}}{1 - T_{EUT}}.$$

Note that if  $\text{RRR} = 0$ , these equations reduce to the Vickers and Elkin DCA equation (which uses "classic"  $T_c$ ).<sup>5</sup>

Scaled by  $B$

$$\begin{aligned} \text{NB}_B[\text{Model}] &= TP_T \cdot \left(1 - \text{RRR} \cdot \frac{\Delta}{B}\right) - FP_T \cdot \frac{H}{B} \\ &= \left(1 - \text{RRR} \cdot \frac{\Delta}{B}\right) \cdot \left(TP_T - FP_T \cdot \left(\frac{T_{EUT}}{1 - T_{EUT}}\right)\right). \end{aligned}$$

Scaled by  $H$

$$\text{NB}_H[\text{Model}] = TP_T \cdot \frac{B - \text{RRR} \cdot \Delta}{H} - FP_T = TP_T \cdot \frac{1 - T_{EUT}}{T_{EUT}} - FP_T.$$

DCA curves are generated by plotting NB of all 3 strategies (eg, "Treat all," "Treat none," and "Model") over all thresholds of interest.

### 2.3 | Derivation of the ERG DCA

Figure 2b depicts the same decision problem from the regret point of view. Utilities of each outcome are now represented in terms of regret, ie, the difference between the utility of the action taken and the utility of the action that, in retrospect, should have been taken.<sup>12-14</sup>

Solving the tree in Figure 2b, we derive the expected regret associated with the prediction model as

$$\text{ERg}[\text{Model}] = \frac{1}{N} \left( \sum_{q_i \geq T} [(1 - p_i)(1 - \text{RRR})(U_4 - U_2)] + \sum_{q_i < T} [p_i(U_1 - U_3)] \right). \quad (6)$$

Just like in EUT case, when the threshold is equal to zero, or one, we have

$$\text{ERg}[\text{Rx none}] = p_D(U_1 - U_3),$$

$$\text{ERg}[\text{Rx All}] = (1 - p_D)(1 - \text{RRR})(U_4 - U_2).$$

We derive the net expected regret difference (NERD) by calculating the expected regret of each choice (eg, "Treat All," "Model") and subtracting it from the expected regret of "Treat None."<sup>7,11,22</sup>

$$\text{NERD}[\text{Model}] = \frac{1}{N} \left( \sum_{q_i \geq T} [(1 - p_i)(U_4 - U_2) - p_i(U_1 - U_3) + p_i \cdot \text{RRR} \cdot (U_4 - U_2)] \right).$$

Using the previously defined  $TP_T$  and  $FP_T$ , as well as  $B$ ,  $H$ , and  $\Delta$ , we can rewrite this as

$$\begin{aligned} \text{NERD}[\text{Model}] &= FP_T \cdot H - TP_T \cdot B + TP_T \cdot \text{RRR} \cdot H \\ &= -(TP_T \cdot (B - \text{RRR} \cdot H) - FP_T \cdot H). \end{aligned}$$

With  $T = 0$  ( $TP_T = p_D$  and  $FP_T = 1 - p_D$ ) in Equation 7, we derive the NERD of the "Treat all" strategy as

$$\text{NERD}[\text{Rx All}] = -(p_D \cdot (B - \text{RRR} \cdot H) - (1 - p_D) \cdot H).$$

The threshold probability is computed as

$$(1-T \cdot (1-RRR)) \cdot (U_4 - U_2) = T \cdot (U_1 - U_3)$$

or

$$T_{ERG} = \frac{H}{B + H - RRR \cdot H}. \quad (7)$$

Note that if  $RRR = 0$ ,  $T_{ERG}$  reduces to the "classic" EUT Pauker and Kassirer (P&K) threshold  $T_c^{1,2}$  (see above).

Scaled by  $B - RRR \cdot H$

$$\begin{aligned} NERD_{B-RRR \cdot H}[\text{Model}] &= -\left( TP_T - FP_T \cdot \frac{H}{B - RRR \cdot H} \right) \\ &= -\left( TP_T - FP_T \cdot \frac{T_{ERG}}{1 - T_{ERG}} \right). \end{aligned}$$

Note that if  $RRR = 0$ , the equations above and below reduce to Vickers and Elkin DCA EUT equation (see also above).<sup>5</sup>

Scaled by  $B$

$$NERD_B[\text{Model}] = -\left( 1 - RRR \cdot \frac{H}{B} \right) \cdot \left( TP_T - FP_T \cdot \frac{T_{ERG}}{1 - T_{ERG}} \right).$$

Scaled by  $H$

$$NERD_H[\text{Model}] = -\left( TP_T \cdot \frac{1 - T_{ERG}}{T_{ERG}} - FP_T \right).$$

### 3 | RESULTS

#### 3.1 | EUT and ERG DCA generate different results

The major difference between EUT and ERG DCA arises because of the definition of the threshold probability when using the classical EUT and utilities expressed via ERG:

$$T_{EUT} = \frac{H}{B + H - RRR \cdot \Delta} \text{ and } T_{ERG} = \frac{H}{B + H - RRR \cdot H}.$$

Both of these thresholds can be connected to the "classical" P&K EUT threshold via

$$\begin{aligned} T_c &= \frac{H}{B + H} = \frac{1}{1 + \frac{B}{H}} \text{ or } B + H = \frac{H}{T_c} \\ T_{EUT} &= \frac{1}{\frac{1}{T_c} - RRR \cdot \frac{\Delta}{H}} \text{ and } T_{ERG} = \frac{1}{\frac{1}{T_c} - RRR} \end{aligned}$$

Again, note that if  $RRR = 0$ , all these thresholds reduce to the "classic" P&K EUT threshold.<sup>1,2</sup>

Extending the threshold model into DCA, we obtain the following:

Scaled by  $B$

$$NB_B[\text{Model}] = TP_T \cdot \left( 1 - RRR \cdot \frac{\Delta}{B} \right) - FP_T \cdot \frac{H}{B},$$

$$NERD_B[\text{Model}] = -\left( TP_T \cdot \left( 1 - RRR \cdot \frac{H}{B} \right) - FP_T \cdot \frac{H}{B} \right),$$

$$NB_B[\text{Rx All}] = p_D \cdot \left( 1 - RRR \cdot \frac{\Delta}{B} \right) - (1 - p_D) \cdot \frac{H}{B},$$

$$NERD_B[\text{Rx All}] = -\left( p_D \cdot \left( 1 - RRR \cdot \frac{H}{B} \right) - (1 - p_D) \cdot \frac{H}{B} \right).$$

Therefore,

$$NB_B[\text{Model}] = -NERD_B[\text{Model}] + TP_T \cdot RRR \cdot \frac{H - \Delta}{B},$$

$$NB_B[\text{Rx All}] = -NERD_B[\text{Rx All}] + p_D \cdot RRR \cdot \frac{H - \Delta}{B}.$$

Scaled by  $H$

$$NB_H[\text{Model}] = TP_T \cdot \left( \frac{B}{H} - RRR \cdot \frac{\Delta}{H} \right) - FP_T,$$

$$NERD_H[\text{Model}] = -\left( TP_T \cdot \left( \frac{B}{H} - RRR \right) - FP_T \right),$$

$$NB_H[\text{Rx All}] = p_D \cdot \left( \frac{B}{H} - RRR \cdot \frac{\Delta}{H} \right) - (1 - p_D),$$

$$NERD_H[\text{Rx All}] = -\left( p_D \cdot \left( \frac{B}{H} - RRR \right) - (1 - p_D) \right).$$

Therefore,

$$NB_H[\text{Rx All}] = -NERD_H[\text{Model}] + p_D \cdot RRR \cdot \frac{H - \Delta}{H}.$$

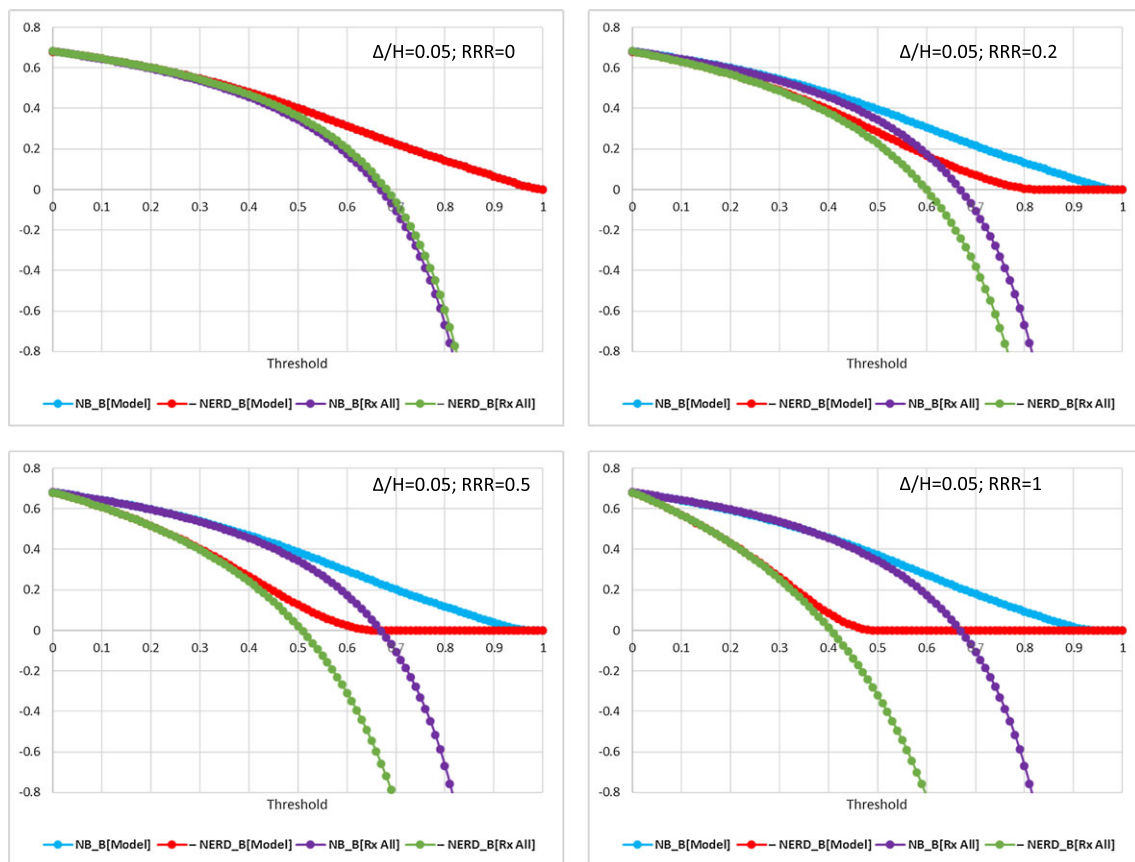
Equations 5 and 7 show that EUT-based DCA and ERG-based DCA differ by threshold definitions ( $T_{EUT}$  vs  $T_{ERG}$ ) and the requirement for specifying  $\Delta$  in the EUT model. As a result, noticeable differences in the evaluation of predictive models will be generated (Figure 3). In this illustrative example, according to the EUT DCA, the use of model to guide a management strategy is almost always best strategy regardless of RRR. However, according to ERG DCA, "Treat None" becomes best strategy with increasing thresholds. Only when  $RRR = 0$ , EUT DCA generates the same results as ERG DCA.

Obviously, different predictive models with different prevalence of disease outcomes, specific values of treatment effects, and disutilities will generate different results than those shown in this example meant for illustration only. We provide an excel file that the reader can use with his/her own model and values for  $B$ ,  $H$ , and  $\Delta$ .

### 4 | DISCUSSION

In this paper, we demonstrate that when treatment effects are included in modeling, different results are generated by EUT and ERG DCA. Under these circumstances, EUT DCA cannot be used to model decisions over all preferences without further knowledge of the specific utilities related to differences between  $U_1$  and  $U_2$  ( $\Delta$ ). This, however, would defy the DCA's original purpose of analyzing decision strategies without requiring the elicitation of patients' preferences.

If the DCA method is to be used, ERG DCA seems to be preferable, which also has the following appeals: (1) it is a mathematically more parsimonious derivation of DCA derived within a coherent regret theory<sup>7</sup>; (2) as a cognitive emotion, regret is widely recognized as one of the key decision making mechanisms enabling a decision maker to experience consequences of decisions both at the emotional (type 1)



**FIGURE 3** Decision curve analysis (DCA) for a model of referral a patient with terminal illness to hospice as a function of the threshold probability  $T$ . Three strategies are considered: “Treat All,” “Treat None: Refer to Hospice” ( $=0$  on  $x$  axis), and “Model: Use Model to Guide Management.” The strategies are equal if they cross each other. The higher the value, the more superior strategy is. NB, net benefit according to expected utility theory (EUT); NERD, net expected regret differences. To enable comparison of EUT DCA and expected ERG DCA, NERD values are presented inversed (ie,  $-NERD$ ). The results clearly show that EUT DCA and ERG DCA generate different results. (Note that  $\Delta/H$  is arbitrary fixed to 0.05; somewhat different results are obtained when  $\Delta/H$  vary.) We used the original patient-level data from the SUPPORT study<sup>23</sup> to create a simpler version of the model concerned with a decision whether to refer a patient to hospice/palliative care in the end-of-life setting. The curves are generated by calculating NB and NERD over all thresholds (from 0 to 1, in increments of 0.01)

and cognitive (type 2) level<sup>3,24,25</sup>; and (3) it is easily and reliably elicited using dual analog visual analog scale (“regret”-meter)<sup>7</sup> or similar scales.<sup>7,26</sup> We should also note that the model we used here is for illustration purposes only—not to advocate the use of this particular model but only to illustrate differences when the model is used within 2 different theoretical frameworks.

Although we advocate using ERG DCA, we are aware of the long tradition of application of EUT in decision sciences, and of the unsettled debate about the superiority of 1 theory over the other. Our main point is that EUT and ERG versions of DCA do generate different results. Although, we cannot possibly settle here the question of superiority of EUT versus ERG (or other non-EUT theories), the larger point we are making is that the decision at which threshold to act closely relates to the question of rational choice.<sup>3,4</sup> The “great rationality debate” has been prominent in nonmedical fields,<sup>27,28</sup> but has only been sporadic in clinical medicine. By highlighting the differences in the results between ERG DCA versus EUT DCA—the latter being widely used method—we hope to stimulate a “rationality debate” in clinical medicine. The practical importance of advancing this debate can be appreciated if we, for example, note that some practice guidelines such as guidelines for colorectal screening are based on

EUT-based modeling<sup>29</sup>; conceivably, different recommendations could have been made if the non-EUT framework were used. Hence, we think that awareness of our findings is of importance to modelers, practicing physicians and policy makers.

## Research Support

This study was supported by the DOD (grant no. W81 XWH 09-2-0175, PI: Djulbegovic).

## REFERENCES

1. Pauker SG, Kassirer J. The threshold approach to clinical decision making. *N Engl J Med*. 1980;302:1109–1117.
2. Pauker SG, Kassirer JP. Therapeutic decision making: a cost benefit analysis. *N Engl J Med*. 1975;293:229–234.
3. Djulbegovic B, Hamm RM, Mayrhofer T, Hozo I, Van den Ende J. Rationality, practice variation and person-centred health policy: a threshold hypothesis. *J Eval Clin Pract*. 2015;21(6):1121–1124. doi: 10.1111/jep.12486. Epub 2015 Dec 7.
4. Djulbegovic B, van den Ende J, Hamm RM, et al. When is rational to order a diagnostic test, or prescribe treatment: the threshold model as an explanation of practice variation. *Eur J Clin Invest*. 2015;45:485–493.

5. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26:565–574.
6. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016;352.
7. Tsalatsanis A, Hozo I, Vickers A, Djulbegovic B. A regret theory approach to decision curve analysis: A novel method for eliciting decision makers' preferences and decision-making. *BMC Med Inform Decis Mak*. 2010;10:51.
8. Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. *JAMA*. 2015;313:409–410.
9. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak*. 2008;8:53.
10. Vickers AJ, Kattan MW, Daniel S. Method for evaluating prediction models that apply the results of randomized trials to individual patients. *Trials*. 2007;8:14.
11. Djulbegovic B, Tsalatsanis A, Hozo I. Determining optimal threshold for statins prescribing: individualization of statins treatment for primary prevention of cardiovascular disease. *J Eval Clin Pract*. 2017;23:241–250.
12. Hozo I, Djulbegovic B. When is diagnostic testing inappropriate or irrational? acceptable regret approach. *Med Decis Making*. 2008;28:540–553.
13. Hozo I, Djulbegovic B. Clarification and corrections of acceptable regret model. *Med Decis Making*. 2009;29:323–324.
14. Bell D. Regret in decision making under uncertainty. *Oper Res*. 1982;30:961–981.
15. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
16. Deeks JJ, Altman DG. Effect measures for meta-analysis of trials with binary outcomes. In: Egger M, Smith GD, Altman DG, eds. *Systematic Reviews in Health Care Meta-analysis in Context*. London: BMJ;2001:313–335.
17. Kent D, Hayward R. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA*. 2007;298:1209–1212.
18. Moons K, Altman D, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.
19. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98:683–690, doi: 10.1136/heartjnl-2011-301246.
20. Sinclair JC, Cook RJ, Guyatt GH, Pauker SG, Cook DJ. When should an effective treatment be used? derivation of the threshold number needed to treat and the minimum event rate for treatment. *J Clin Epidemiol*. 2001;54:253–262.
21. Djulbegovic B, Hozo I, Lyman G. Linking evidence-based medicine therapeutic summary measures to clinical decision analysis. *MedGenMed* 2000. Available at [http://www.medscape.com/viewarticle/403613\\_1](http://www.medscape.com/viewarticle/403613_1).
22. Tsalatsanis A, Barnes LE, Hozo I, Djulbegovic B. Extensions to regret-based decision curve analysis: an application to hospice referral for terminal patients. *BMC Med Inform Decis Mak*. 2011;11:77.
23. Knaus W, Harrell F, Lynn J, et al. The SUPPORT prognostic model: objective estimates of survival for seriously ill hospitalized adults. *Ann Intern Med*. 1995;122:191–203.
24. Kahneman D. *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux;2011.
25. Djulbegovic B, Beckstead JW, Elqayam S, et al. Evaluation of physicians' cognitive styles. *Med Decis Making*. 2014;34:627–637.
26. Djulbegovic M, Beckstead J, Elqayam S, et al. Thinking styles and regret in physicians. *PLoS One*. 2015;10: e0134038.
27. Stanovich KE. *Rationality and the Reflective Mind*. Oxford: Oxford University Press; 2011.
28. Schwartz B. What is rationality? *J Mktg Behav*. 2015;1:175–185.
29. Knudsen AB, Zauber AG, Rutter CM, et al. Estimation of benefits, burden, and harms of colorectal cancer screening strategies: modeling study for the US preventive services task force. *JAMA*. 2016;315:2595–2609.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Hozo I, Tsalatsanis A, Djulbegovic B. Expected utility versus expected regret theory versions of decision curve analysis do generate different results when treatment effects are taken into account. *J Eval Clin Pract*. 2018;24:65–71. <https://doi.org/10.1111/jep.12676>