Will Insistence on Practicing Medicine According to Expected Utility Theory Lead to an Increase in Diagnostic Testing? Reply to DeKay's Commentary: Physicians' Anticipated Regret and Diagnostic Testing

Iztok Hozo, PhD, Benjamin Djulbegovic, MD, PhD

EXPECTED UTILITY LIABILITY V. ACCEPTABLE REGRET MODEL

Michael DeKay[1] contends that patients are better off if physicians act based on expected utility theory (EUT). This is, of course, a several decades-old debate. Our article is an attempt to contribute to this debate by providing a descriptive (and not normative) account for some observed decision-making behavior.[2] However, one of our key messages is that not behaving according to the EUT rationality criterion does not make behavior irrational.[3] In the same fashion, consideration of testing from physicians' point of view, although raising important ethical challenges, may not necessary jeopardize patients' best interests.[3]

It is undeniable fact that physicians do *not* act according to EUT. Consider an example of treatment and diagnosis of pulmonary embolism (PE) we discussed in our article.[2] According to EUT, if no diagnostic test is taken into consideration, physicians should treat patients with suspected PE as long as probability of pPE > 1.62%. Or, if a spiral computed tomography (CT) scan is available, they should order it for any pPE > 0.07%. No physician, even those of us who are staunch believers in the EUT rationality criterion, does it. Most experts and guidelines panels recommend that PE should be excluded in patients with a probability of PE lower than 5%, the value that is 72 times higher than the EUT threshold of 0.07%![4,5] DeKay[1] offers the EUT malpractice liability model as an alternative, arguing that defensive testing (presumably aimed at reducing physicians' regret about being wrong) harms patients.[6] The model, however, stops short of calculating the exact probabilities of, say, PE, at which DeKay would recommend treating patients with PE or recommend ordering a CT angiogram. Only when one considers actual outcomes (thresholds) can he or she get better appreciation of the usefulness and validity of models. Below, we provide our attempt to illustrate the results of different thresholds at which physicians can be expected to order a diagnostic test according to EUT, EUT-liability, and acceptable regret model. In the accompanying piece, we also provide corrections of our original model as well as further clarifications under which circumstances the acceptable regret model can explain both overtesting and undertesting (when EUT is used as a criterion for diagnostic over- and undertesting).[7]

We first note that DeKay[1] does, in fact, acknowledge that "it may be reasonable to ignore small, clinically irrelevant differences between the expectations of treatment options." What constitutes "small, clinically irrelevant differences" and how to operationalize them is, however, a crucial issue. This is, in fact, the basis of acceptable regret (R_o), which we defined as a loss in utility when undertaking a wrong decision that will not be particularly burdensome to the decision maker.[2,8] The conceptual advantage of our model is that is derived from regret—that is, one of the basic

From the Department of Mathematics, Indiana University Northwest, Gary, Indiana (IH) and University of South Florida, Clinical Translational Science Institute, Center for Evidence-Based Medicine and Health Outcomes, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida (BD).

Address correspondence to Benjamin Djulbegovic, MD, PhD, USF Health Clinical Research, 12901 Bruce B. Downs Boulevard, MDC02, Tampa, FL 33612; e-mail: bdjulbeg@health.usf.edu.

DOI: 10.1177/0272989X09334370

Table 1 Benefit and Harms of Treatment of Pulmonary Embolism with Anticoagulants[2]

Treatment	Mortality without Treatment, %	Mortality with Treatment, %	Efficacy, %	Absolute Death Reduction, %	Harms of Treatment (Death), %	Benefit/Harms
Heparin → warfarin (''typical'' case)	25	2.5	90	22.5	0.37	60.8
Worst-case scenario	25	7.5	70	17.5	5	3.5

Test Characteristics for Diagnosis of Pulmonary Embolism

Test	Sensitivity, %	Specificity, %	Harms (Death Caused by Testing), %
Pulmonary angiogram	99[a]	99[a]	0.05
Computed tomography angiogram	97	99	0

a. Gold standard test, which does not have 100% sensitivity and specificity due to operator error.

psychological tenets of decision making and does not require the assumption of ''stolen utility.''[1]

TEST THRESHOLDS ACCORDING TO EUT, EUT-LIABILITY, AND ACCEPTABLE REGRET MODEL

By illustrating specific testing thresholds, which can be derived from each model, we are in better position to argue which model is more acceptable. In the EUT framework, the test threshold (p_{tt}) represents the probability of disease at which we are indifferent between ordering a diagnostic test and withholding treatment.[9] Test threshold depends on the treatment's net benefits (B) and harms (H), test operating characteristics, and risk of the test (H_{te}; see Table 1). DeKay and Asch's[10] malpractice liability EUT model incorporates additional variables in the classic threshold model: L_{rx}, a physician's expected liability for providing unnecessary treatment; L_{nrx}, a physician's expected liability for failing to provide necessary treatment; LR_{rx}, a reduction in a physician's expected liability for providing unnecessary treatment afforded by obtaining a diagnostic test; and LR_{nrx}, a reduction in a physician's expected liability for failing to provide necessary treatment afforded by obtaining a diagnostic test.

DeKay and Asch[10] do not provide any clinical data on liability variables, but they assume that $L_{nrx} \geq LR_{nrx} \geq 0$ and $L_{rx} \geq LR_{rx} \geq 0$. To illustrate the main point of our analysis, we will assume two extreme scenarios in DeKay and Asch's[6] malpractice liability EUT model: 1) maximum liability (i.e., L_{nrx} ($= B = 17.5\%$) and L_{rx} ($= H = 5\%$)) and zero protection by ordering tests (both LR_{nrx} and LR_{rx} are assumed to be 0) and 2) maximum liability reduction (i.e., $L_{nrx} = B$ and $L_{rx} = 0$) and maximum protection by ordering diagnostic tests ($LR_{nrx} = B$, $LR_{rx} = 0$). Finally, we will assume R_o of 1% of net benefits—that is, we can tolerate a loss of 1% of death reduction ($= 0$ .175%) in case we were wrong to order a pulmonary angiogram.

Calculations show that testing thresholds (in case of ordering a pulmonary angiogram) according to classic EUT = 0.58%, DeKay and Asch's model = 0.43% (maximum liability) and 0.29% (maximum liability protection), and the acceptable regret model = 60%. Incidentally, treatment thresholds for EUT = 95.6%, DeKay and Asch's model = 96.1% (maximum liability) and 95.61% (maximum protection), and the acceptable regret model = 96.5%. (An Excel sheet to help calculate thresholds is available from the authors upon request.) While conceptually EUT and the liability model may be different, in actuality both models will produce similar recommendations that defy the reality of clinical practice. In contrast, the acceptable regret model generates thresholds that are more reflective of the way how medicine is practiced. In fact, the differences between acceptable regret and other models are rather dramatic (differences are even more pronounced for harmless tests and tests that are less accurate). It is, however, important to realize that while ordering of a diagnostic test (e.g., pulmonary angiogram in our case) at the pretest probability, which is less than the acceptable regret threshold is appropriate (from a descriptive point of view), it is not required action on a part of the physician (i.e., the acceptable regret threshold does not represent a normative maxim, which physicians ought to follow). The physician is actually free to act within his or her range of the acceptable regret threshold (i.e., 0%–60% in our example) as long he or she desires to keep regret of a potentially wrong decision acceptable. This may mean that some physicians may act according to EUT; in fact, one can calculate a zone within which readiness to tolerate a wrong decision is still compatible within EUT.[11] Nevertheless, the observations from everyday practice strongly indicate that physicians routinely violate EUT but often act according to acceptable regret theory.

After arguing for more than 2 decades that physicians are not behaving appropriately, we think it is time to ask, ''Is it really the case that all physicians are irrational when they do not order diagnostic tests according to EUT?'' Are people stupid, or our models are not accurate? In the PE example, we are not aware of a single physician who would order an angiogram at a probability of disease that is less than 1%, but many who would do it at the probability around 60%.

We believe that the concept of acceptable regret (R_o) can successfully explain some of dramatic differences between what physicians do v. what they should do. Which behavior is more rational will, undoubtedly, continue to be debated. However, we do want to note one interesting phenomenon: because in today's practice, benefits of approved treatments vastly outweigh their harms, and because most diagnostic tests are perceived to be harmless with decent sensitivity and specificity, the testing and treatment thresholds will predictably be low for the majority of tests and interventions employed in the contemporary practice. This means that insisting on practicing medicine according to EUT would lead to a further increase in the use of diagnostic and treatment interventions.* Most commentators, however, believe that further increase in the use of tests and therapeutic interventions cannot be considered rational. We hope that that acceptable regret model may shed some light on this crucial issue in the practice of medicine.[2,7,12,13]

REFERENCES

1. DeKay ML. Physicians' anticipated regret and diagnostic testing: comment on Hozo & Djulbegovic, 2008. Med Decis Making. 2009;29:**[IN PRESS]**.
2. Hozo I, Djulbegovic B. When is diagnostic testing inappropriate or irrational? Acceptable regret approach. Med Decis Making. 2008;28:540–53.
3. Kahneman D. New challenges to the rationality assumption. J Inst Theor Econ. 1994;150:18–36.
4. Roy P-M, Colombet I, Durieux P, Chatellier G, Sors H, Meyer G. Systematic review and meta-analysis of strategies for the diagnosis of suspected pulmonary embolism. BMJ. 2005;331:259.
5. Hull RD. Diagnosing pulmonary embolism with improved certainty and simplicity. JAMA. 2006;295:213–5.
6. DeKay ML, Asch DA. Is the defensive use of diagnostic tests good for patients, or bad? Med Decis Making. 1998;18:19–28.
7. Hozo I, Djulbegovic B. Clarifications and corrections of acceptable regret model. Med Decis Making. 2009;29**[IN PRESS]**.
8. Djulbegovic B, Hozo I, Schwartz A, McMasters KM. Acceptable regret in medical decision making. Med Hypotheses. 1999;53:253–9.
9. Pauker SG, Kassirer J. The threshold approach to clinical decision making. N Engl J Med. 1980;302:1109–17.
10. DeKay ML, Asch DA. Offensive testing: the balancing act, the evil twin, and the pure play. Med Decis Making. 1998;18:35–6.
11. Djulbegovic B, Hozo I. When should potentially false research findings be considered acceptable? PLoS Med. 2007;4(2):e26.
12. Djulbegovic B, Hozo I, Schwartz A, McMasters K. Acceptable regret in medical decision making. Med Hypotheses. 1999;53:253–9.
13. Hozo I, Schell MJ, Djulbegovic B. Decision-making when data and inferences are not conclusive: risk-benefit and acceptable regret approach. Semin Hematol. 2008;45:150–9.

*Considerations of costs may change this impression. This does not mean that practicing according to acceptable regret would reduce diagnostic testings. However, the acceptable regret model does imply that the efforts to reduce costs by focusing on reducing the amount of diagnostic testing will likely not succeed.

Clarifications and Corrections of Acceptable Regret Model

1. WHEN THE TEST HAS AN ASSOCIATED HARM

We thank Dr. DeKay for pointing out an error in this section of our manuscript.[1] Indeed, when the error is corrected, the expected regret thresholds are identical to the expected utility thresholds.[1,2] In fact, as also pointed out in our article, when regret is a linear function, the regret theory and expected utility theory produce the same results.[3,4] The conceptual advantage of our model is that it is derived from regret (i.e., one of basic psychological tenets of decision making).

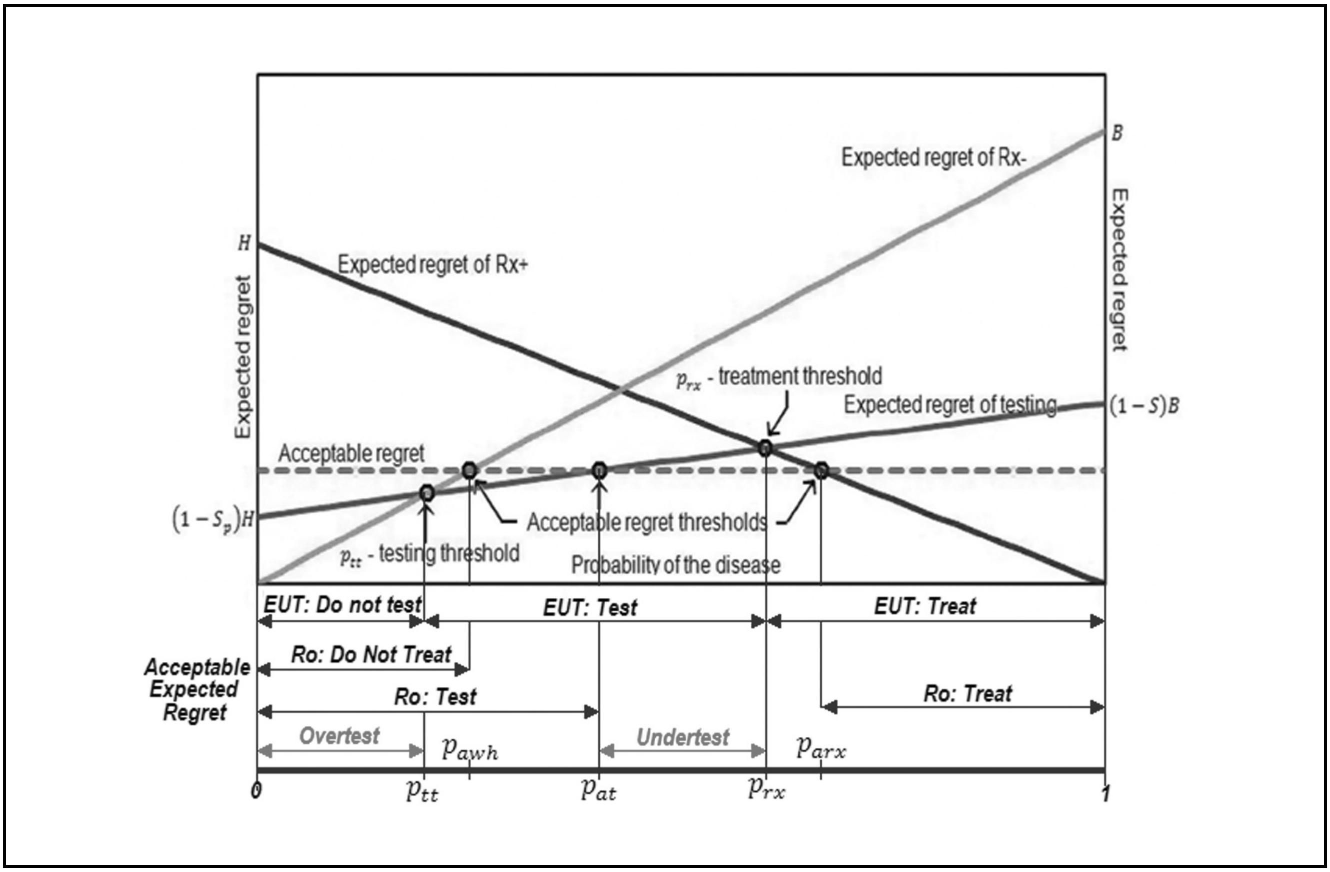

Figure 1 Illustration how acceptable regret can explain both overtesting and undertesting (when appropriateness of testing is defined by expected utility theory [EUT]). If the probability of the disease is smaller than the acceptable expected regret theory (AERT) testing threshold, p_{at}, the AERT would lead to test ordering, whereas according to EUT, we should not test. This, in turn, leads to overtesting. If the probability of the disease is larger than the AERT testing threshold, p_{at}, according to AERT, we would be reluctant to order the test, whereas according to EUT, we should test. As a consequence, this would lead to undertesting. We think that overtesting typically occurs in the "rule out worst-case scenarios" in which physicians cannot afford to miss a particular diagnosis. Once a serious diagnostic possibility enters the physician's mind, every patient with chest pain or shortness of breath gets a computed tomography (CT) angiogram to rule out pulmonary embolism (PE), every patient with headache receives a CT of the head to rule out brain tumor, every patient with "incidentaloma" (incidental and unexpected finding of a mass on imaging studies performed for different reasons) gets (continued)

(continued)
a biopsy, and so on. We think undertesting typically occurs when ordering a diagnostic test is perceived as not needed (i.e., consciously or subconsciously is felt to be risky or associated with unacceptable level of regret). Hence, patients with atypical chest pain will not get a CT angiogram to rule out PE, patients with headache do not get a CT of the head, patients with ''incidentaloma'' will not get a biopsy, and so on. P_{tt}, testing threshold according to EUT; P_{rx}, treatment threshold according to EUT; P_{at}, testing threshold according to acceptable regret (R_o) model; p_{awh}, threshold probability below which treatment can be withheld without experiencing regret (if decision was wrong). For details, see Hozo and Djulbegovic.[3]

The corrected expression is

$$E_{rg}[T] = S \cdot p \cdot H_{te} + (1 - S_p) \cdot (1 - p) \cdot (H + H_{te}) + (1 - S) \cdot p \cdot (B + H_{te}) + S_p \cdot (1 - p) \cdot H_{te}$$

This, in turn, requires corrected versions of formulas (7), (8), and (9) in our original manuscript as[3]

$$p_{rx} = \frac{S_p H - H_{te}}{S_p H + (1 - S)B} = \frac{1 - \frac{1}{S_p}\frac{H_{te}}{H}}{1 + (LR-)\frac{B}{H}} \tag{7}$$

$$p_{tt} = \frac{(1 - S_p)H + H_{te}}{(1 - S_p)H + S \cdot B} = \frac{1 + \frac{1}{(1 - S_p)}\frac{H_{te}}{H}}{1 + (LR+)\frac{B}{H}} \tag{8}$$

$$p_{at} = \frac{(1 - S_p)H - R_o + H_{te}}{(1 - S_p)H - (1 - S)B} = \frac{1 - \frac{1}{(1 - S_p)}(R_o - H_{te})}{1 - \left(\frac{1}{1 - S_p} - (LR+)\right)\frac{B}{H}} \tag{9}$$

This means that the ''Test'' column in the ''Acceptable regret threshold'' part of Table 2 should be corrected. Only 2 numbers in that column of Table 2 should be changed: 0.00% instead of 20.14% and 0.00% instead of 17.75% (fourth from the bottom and last).* Although unfortunate, none of this invalidates the concepts proposed in our manuscript.

2. ACCEPTABLE REGRET CAN EXPLAIN BOTH OVERTESTING AND UNDERTESTING

DeKay correctly points out that the acceptable regret threshold (*pat*) should be understood as the upper bound for diagnostic testing (whenever $(1 - S)B \geq (1 - Sp)H$). Thus, testing is always acceptable if the probability of disease is lower than *pat*. We also thank Dr. DeKay for correcting our interpretation of overtesting and undertesting. Figure 1 shows a correct interpretation of how the acceptable regret threshold model can explain both overtesting and undertesting.

REFERENCES

1. DeKay ML. Physicians' anticipated regret and diagnostic testing: comment on Hozo & Djulbegovic, 2008. Med Decis Making. 2009;29:**[IN PRESS]**.

2. Djulbegovic B, Hozo I, Schwartz A, McMasters KM. Acceptable regret in medical decision making. Med Hypotheses. 1999;53:253–9.

3. Hozo I, Djulbegovic B. When is diagnostic testing inappropriate or irrational? Acceptable regret approach. Med Decis Making. 2008;28:540–53.

4. Loomes G, Sugden R. Regret theory: an alternative theory of rational choice. Econ J. 1982;92:805–24.