Testing the descriptive performance of the rank-dependent utility in the domain of health profiles[1]

[Accepted in the *Spanish Economic Review*, March, 2001]

José Mª Abellán Perpiñán*
Department of Applied Economics
University of Murcia
Murcia (Spain).
E-mail: dionisos@um. Fax-number: +34 968 363745.

José Luis Pinto Prades
Department of Economics
Center for Health Economics
Pompeu Fabra University
Barcelona (Spain).

Key words: Expected utility; health outcomes; medical decision-making; QALYs; rank-dependent utility

JEL: D8, I1.

* Correspondence to: José María Abellán Perpiñán, Department of Applied Economics, University of Murcia, Campus of Espinardo, 30100 Murcia, Spain.

---

[1] We thank a co-editor and two anonymous referees for their helpful comments. Support from research grant SEC2000-1087 by Ministerio de Ciencia y Tecnología is gratefully acknowledged.

Abstract

Expected utility theory (EUT) has been challenged as a descriptive theory in many contexts. The medical decision analysis context is not an exception. Several researchers have suggested that rank dependent utility theory (RDUT) may accurately describe how people evaluate alternative medical treatments. Recent research in this domain has addressed a relevant feature of RDU models - probability weighting - but to date no direct test of this theory has been made. This paper provides a test of the main axiomatic difference between EUT and RDUT when health profiles are used as outcomes of risky treatments. Overall, EU best described the data. However, evidence on the editing and cancellation operation hypothesized in Prospect Theory and Cumulative Prospect Theory was apparent in our study. We found that RDU outperformed EU in the presentation of the risky treatment pairs in which the common outcome was not obvious. The influence of framing effects on the performance of RDU and their importance as a topic for future research is discussed.

1. Introduction

During the past decades, experimental evidence involving monetary outcomes has challenged expected utility theory (EUT) as a descriptive theory of decision under risk. Numerous paradoxes have displayed systematic violations of the axioms underlying EUT, mainly with respect to its crucial axiom, independence. As it is well-known, the independence axiom constrains EU to be "linear in probabilities", however, many empirical studies, beginning with the seminal test of Preston and Baratta (1948), have systematically suggested that individuals do not treat probabilities in a linear way.

As solutions to the demonstrated violations of EUT, new theories of decision under risk have been proposed. Among the new axiomatic nonexpected utility theories, rank-dependent utility (RDUT), first introduced by Quiggin (1982), emerges as the most popular alternative to classical expected utility. The key idea of this theory is that people transform probabilities into nonlinear decision weights, through a cumulative probability weighting function, in a rank-dependent manner. Cumulative prospect theory (CPT), introduced by Tversky and Kahneman (1992), applies the Quiggin's rank-dependent method for transforming probabilities separately to gain and losses. In this way, CPT may explain nonlinearities in probabilities and gain-loss asymmetry altogether on the basis of rank and sign-dependence. CPT coincides with RDU theory when all outcomes are gains.

When comparing RDU with other nonexpected utility theories, empirical evidence strongly favors RDU. In general, most past empirical studies can be classified

into one or more of the three following methodological approaches to test RDU. Many studies have tested a variety of competing nonexpected utility theories through their predictions on the shape of indifference curves in the probability triangle (Camerer 1989, 1992, 1995). Other studies have fitted both the utility function and the probability weighting function to individuals (Lattimore, Baker and Witte 1992, Tversky and Kahneman 1992, Wu and Gonzalez 1996). Finally, several researchers have developed nonparametric methods to elicit the probability weighting function (Currim and Sarin 1989, Wu and Gonzalez 1998) and, even, the utility function and the probability weighting function succesively (Abdellaoui 1999).

However, previous tests have in common that they tested a prominent feature of RDU theory, probability weighting, but they do not tested RDU theory itself. RDU theory is an axiomatic decision theory and, as Wakker et al. (1994) assert, "axiomatizations describe the testable implications of a model of choice" (p. 197). Thus a direct test of a decision theory can only be achieved by testing those axioms that give empirical content to the theory. That is the "yardstick" to accept or reject any theory competing with EU. Of course, this does not mean that we should give up to other approaches, e.g. goodness-of-fit tests. However, we should be aware that such approaches will be incomplete unless we investigate the descriptive validity of the axioms of the theory.

With respect to RDU theory, the axiomatic yardstick was formally provided by Wakker (1996) who showed that the exact demarcation between EU and RDU is the difference between the independence axiom and a weaker form of independence called comonotonic independence. Other axiomatic differences between both theories are

nonessential. Thus a test between independence and comonotonic independence can be used as a test of the descriptive validity of EU as compared with RDU in the context of risk (or uncertainty).

Previous attempts to test comonotonic independence in the monetary domain obtained negative findings for RDU (Wakker et al. 1994, Fenemma and Wakker 1996). Weber and Kirsner (1997) found a positive result when judgments on prices were required but not when individuals made choices between gambles. In our opinion, despite these negative findings the theory can not be rejected. No theory can probably explain all risky choice behavior in all decision-making contexts, but different alternative theories can be of descriptive value in different specific domains of decision making. As a consequence, an interesting topic of research is to identify domains in which people behave according to some specific theory. This paper intends to contribute to this identification task by testing the descriptive validity of EU and RDU in the specific domain of the health outcomes.

Recent research in this context have explored the role of probability weighting in the elicitation of health states utilities (Wakker and Stiggelbout 1995, Bleichrodt 1996, Bleichrodt 1999, Bleichrodt et al. 1999a, Bleichrodt and Pinto 1999) and even how incorporate RDU and CPT into the medical decision analysis (Bleichrodt and Quiggin 1997, Bleichrodt et al. 1999b, Bayoumi and Redelmeier 2000) but to date, to our knowledge, no direct test between the validity of both EUT and RDUT for health outcomes has been performed. Hence this study performs a direct comparison of the theories by testing violations of independence versus comonotonic independence in the medical decision making domain.

The structure of the paper is as follows. Section 2 briefly explains the difference between EU and RDU through the key role that independence and comonotonic independence axioms play in each theory. Section 3 describes the experimental design used to test RDU against EU. The results are described in Section 4. Discussion in Section 5 closed the paper.

2. The RDU theory and the comonotonic independence principle in medical decision making

The normative theory by excellence in decision making under risk has been (and is) expected utility. The mathematical expression of EU is well known:

$$EU=\sum_{i=1}^{n} p(x_i)u(x_i) \qquad [1]$$

Eq. (1) implies that each outcome is weighted linearly by its probability of occurrence. However, empirical findings tend to show that this may not be so. Instead, it seems to suggest that the relevance of outcomes is not proportional to the associated probabilities. Obviously, such a kind of preferences is not consistent with the EU rule described by Eq. (1).

This is one of the reasons that has lead some authors to suggest that EU is not a good descriptive theory of decision making under risk. RDU accommodates EU deviations substituting the cumulative probabilities by decision weights. For example, let us assume that somebody is diagnosed with a certain illness that causes migraine quite frequently (say 8 days per month). He is offered treatment A and its associated outcomes and probabilities are (0.1, complete recovery; 0.7, 6 days of migraine per month; 0.2, no improvement). To evaluate the utility of this treatment under RDU theory we have to follow five steps:

1. Outcomes are rank-ordered from least to most preferred, i.e., no improvement $\prec$ 6 days of migraine $\prec$ complete recovery, where the individual preference relation $\prec$ means "less preferred to".

2. Both the cumulative probability of reaching each outcome or any other more preferred and the cumulative probability of reaching all other outcomes more preferred are calculated:

Probability of reaching at least no improvement = 1

Probability of reaching all other outcomes better than no improvement = 0.8

Probability of reaching at least 6 days of migraine per month = 0.8

Probability of reaching all other outcomes better than 6 days of migraine per month= 0.1

Probability of reaching at least complete recovery = 0.1

Probability of reaching all outcomes better than complete recovery = 0

3. Each of the cumulative probabilities is transformed by a weighting function w:

$1 \rightarrow w(1)$, $0.8 \rightarrow w(0.8)$, $0.1 \rightarrow w(0.1)$, $0 \rightarrow w(0)$, where $w(0)=0$ and $w(1)=1$.

4. The decision weight φ associated with each outcome is calculated as the difference between the transformed cumulative probability of reaching at least the outcome and the transformed cumulative probability of reaching all other outcomes more preferred:

φ(no improvement) = $1 - w(0.8)$

φ(6 days of migraine per month) = $w(0.8) - w(0.1)$

φ(complete recovery) = $w(0.1)$

5. Finally, the RDU of the gamble (0.2, no improvement; 0.7, 6 day of migraine per month; 0.1, complete recovery) is calculated as follows:

[1 – w(0.8)]u(no improvement) + [w(0.8) – w(0.1)]u(6 days migraine per month) + w(0.1)u(complete recovery) [2]

In general, the RDU of a gamble yielding outcome q_i with probability p_i (i=1,2,..., n), where the outcomes are rank-ordered from least to most preferred, $q_1 \prec q_2 \prec ... \prec q_n$, can be represented by:

$$\text{RDU}=\sum_{i=1}^{n} \varphi_i u(q_i) \qquad [3]$$

where the decision weights φ_i are defined by:

$$\varphi_i = w\left(\sum_{i=j}^{n} p_i\right) - w\left(\sum_{i=j+1}^{n} p_i\right) \qquad [4]$$

As it has been shown, RDUT assumes that the weight attached to each outcome in a gamble is dependent on the ranking of the outcomes, in order of preference, whereas in EUT the weights of the outcomes (the probabilities) are independent on the rank-order. The difference between the two theories is based on the axioms that underlie each theory. Whereas EU is based on the independence axiom, RDU is based on the comonotonic independence axiom.

The independence axiom implies that if we have two gambles A and B with a common outcome and A $\succ$ B, then preferences will not be affected if a common outcome is replaced by another common outcome. If preferences change (B $\succ$ A)

independence is violated and EU fails as a descriptive theory of choice under risk. The comonotonic independence axiom implies that if we have two gambles A and B with a common outcome and A $\succ$ B, then preferences may be affected if a common outcome is replaced by another common outcome so long as the ranking of the common outcome changes. If preferences change (B $\succ$ A) comonotonic independence is not violated. However, if the new common outcome does not change the rank-order of the outcomes in the gambles, then preferences should remain still constant. If preferences change comonotonic independence is indeed violated and RDU fails to describe observed behavior. When the rank-order changes we will talk of a noncomonotonic change and the reverse.

Now we describe the implications of both comonotonic and noncomonotonic independence with the migraine example used above. Let us assume that there is an alternative treatment (B) that gives rise to next gamble where outcomes are rank-ordered from worst to best:

(0.2, no improvement; 0.7, 5 days of migraine; 0.1, 1 day of migraine)

and let us assume that the patient prefers gamble A (the former treatment) to gamble B.

Let us now change the 20% of no improvement in both gambles by a 20% chance of 4 days of migraine. The new gambles would then be:

A’ = (0.2, 4 days of migraine; 0.7, 6 days; 0.1, complete recovery)

B’ = (0.2, 4 days of migraine; 0.7, 5 days; 0.1, 1 day of migraine)

Under EU if treatment A $\succ$ treatment B then A' $\succ$ B', however under RDU it might happen that the individual prefers A to B when the common outcome is "no improvement " but she prefers treatment B' to A' when the common outcome is "4 days of migraine". The change of the common outcome from "no improvement" into "4 days of migraine" implies that the common outcome is no longer the worst ranked outcome. There has thus been a noncomonotonic change in the common outcome. Under EUT, independence should still hold, and individual preference should still remain constant. However, under RDU, independence holds so long as the ranking of the common outcome remains the same. If the rank-order changes preferences are then allowed to vary. Thus, under RDU, the subject who initially preferred treatment A is now allowed to prefer treatment B'.

Not many tests have been conducted comparing independence and comonotonic independence. As we have already mentioned, most of the studies have confirmed the hypothesis that probabilities are transformed using both parametric and nonparametric methods. However, the more basic and pure test to compare the descriptive validity of RDU against EU - the direct comparison between comonotonic independence and independence - has not been conducted until now in the medical decision making context. We will then attempt to answer the question: what axiom best describe choices between medical interventions, independence or comonotonic independence?

3. The experiment

- Stimulus design

We constructed several choices between pairs of (risky) medical treatments. Two different contexts were used depending on the severity of the health problem

described to the participants in the experiment, cancer or dermatitis. Each medical treatment was presented as a gamble. Outcomes were described as years in a certain health state until the patient returns to normal health. Two different health states were used both for cancer and dermatitis (Appendix 1). The probabilities were contingent upon the kind of the disease experienced. Three different kind of diseases were possible in each context: spread epidermoid cancer, non-spread epidermoid cancer, and esophagic adenocarcinoma (in the cancer context), and irritative contact dermatitis, alergic contact dermatitis, and photodermatitis (in the dermatitis context). Each pair of treatments in a choice problem always showed, under some of the three possible diseases, a common outcome.

[Insert Figure 1 about here]

We further used four different gamble displays in order to control possible *presentation effects*. There exists ample evidence in the decision analysis literature about the influence of the display format on preferences (Tversky and Kahneman 1981, 1986). Figure 1 presents the display of one pair of gambles under the four formats. For example, if the subject has allergic dermatitis and she chooses treatment 1 she will need 8 years to be cured. In the meantime she will be in health state A worse than full health.

We thought that the inclusion of different experimental conditions, i.e., two health contexts, two health states and four displays, would allow us to examine the consistency of the performance of RDU. For example, if RDU would work in a context but would fail in the other we should then be cautious to value the role of RDU for medical decision analysis.

[Insert Figure 2 about here]

We constructed 42 pairs of gambles for each context, but only 32 pairs were relevant in order to test RDU against EU. These 32 pairs were distributed across eight sets of four pairs each: four sets under condition "health state A" and four sets under condition "health state B". Each pair of gambles consisted of a "safer" gamble S and a "riskier" gamble R with similar expected value but different variance.

The four pairs within each set were identical except for the common outcome. All sets provided one change of the common outcome that did not affect the rank-ordering of the outcomes and two changes of the common outcome that affected the rank-ordering. Thus, each set offered one test of comonotonic independence and two tests of noncomonotonic independence. As an example, consider set 1 in Figure 2. As the change of the common outcome 10 into 8 does not affect the rank-order of the outcomes, the choices in these two first pairs of the set provide a test of comonotonic independence. In contrast, the next two changes of the common outcome ($8 \rightarrow 6$; $6 \rightarrow 4$) do affect the rank ordering of the outcomes, providing two tests of noncomonotonic independence.

Four pairs were replicated in the questionnaire. The replications were added to provide some indication of the degree of consistency of the subjects. The remaining pairs up to 42 were those that Wakker et al. (1994) named filler pairs, i.e., pairs with clearly distinct expected values to motivate subjects to examine the choice alternatives carefully.

- *Subjects and procedure*

Prior to the actual experiment was conducted, the questionnaire was checked in various pilot sessions with members of the academic staff at the University of Murcia. Seventy-two economic students of this University participated in the experiment. Subjects were randomly assigned to one of two contexts. Thirty-five subjects were assigned to the first class of disease (cancer) and thirty-seven to the second (derma). In the same way, they were randomly assigned to one of the four displays. The experiment was carried out in two sessions. In first session participants received information about the experiment, instructions were read aloud to them and several practice trials were made to avoid *learning effects*. In the second session students were given the questionnaire with the choice questions to test EU versus RDU. Subjects were told to imagine having some symptoms common to a group of three diseases. Subjects had to choose the preferred treatment from each of 42 pairs of treatments. The order of the questions was varied to avoid *order effects*. After all choices had been made, the subjects were debriefed and given credit for participating.

- *Hypothesis and statistical analysis*

We expect that RDU would perform better than EU in the specific context of medical decision-making. The theoretical method chosen in our experiment to test the performance of RDU relative to EU is the direct comparison between the performance of comonotonic independence and independence. As we have described above, each set of risky treatments presented to the respondents offered one test of comonotonic independence and two tests of noncomonotonic independence. Overall, three tests of independence by set. Thus if RDU performs better than EU the main hypothesis of the experiment predicts that the number of comonotonic independence violations will be less than one-third of the violations of independence. In other words, that violations of

independence are more likely for noncomonotonic changes of the common outcome as for comonotonic ones. The null hypothesis assumes, on the contrary, that the number of violations of comonotonic independence is one-third of the number of violations of independence. We will further define three sub-hypothesis, namely, the main hypothesis should be satisfied for different health contexts (sub-hypothesis 1), different health states (sub-hypothesis 2), and different presentations of the choices (sub-hypothesis 3).

The main hypothesis will be then confirmed if the proportion of violations of comonotonic independence denoted by Vci/Vind < 1/3, where Vci and Vind represent the observed number of violations of comonotonic independence and independence respectively. Statistical significance will be tested with both parametric (one-sample Z-test) and non-parametric (binomial test) techniques. All tests will be one-side with $\alpha = 0.05$. We will do the same with sub-hypothesis 1-3.

We will further investigate the performance of RDU and EU across the different experimental conditions. For example, if sub-hypothesis 1 holds then RDU will outperform EU in both health contexts. We will then test if RDU works better in one context than in other comparing the proportion Vci/Vind observed in the two contexts. Statistical significance would be tested with both parametric (two-sample Z test) and non-parametric (two-sample binomial test) techniques. Otherwise, if sub-hypothesis 1 does not hold we will then test if EU works better in one context than in other. We will then compare the rate of violation of independence reached in both contexts. For example, in the case of the cancer sample, we have that for 35 respondents, 8 choice sets (4 sets with health state A and 4 sets more with state B), and 3 test of independence per set, there are altogether 35x8x3 = 840 tests of independence. In consequence, in the

cancer sample we know that at best 840 violations can occur. Thus, if 50 violations are observed the rate of independence violation will be 50/840 = 5.95%. Differences will be tested with both parametric (two-sample t test) and non-parametric (Mann-Whitney U-test) techniques. Similar procedures will be used with the remaining experimental conditions.

4. Results

The left half of Table 1 lists the number of violations of comonotonic independence and EU-independence. The right half displays both the proportions of violations of comonotonic independence observed (Pci obsd.) and the difference with respect to the proportion predicted by the null hypothesis (1/3). The table shows that the main hypothesis of the experiment does not hold. In the same way, no sub-hypothesis holds. We only found partial evidence of rank-dependence for the collapsed tree presentation ($p < 0.05$). This latter result, although very modest, has not been reported before in those studies that have tested comonotonic independence with monetary outcomes (Wakker et al. 1994, Fennema and Wakker 1996, Weber and Kirsner 1997).

We do not think that this finding can be explained by inter-sample differences given that the subjects that used the collapsed tree format came from the same population (University of Murcia students). Apart from that, the different displays were assigned randomly. Nevertheless, the result has to be put into perspective since the sample size that used this presentation was small ($n = 18$). Thus this positive evidence of comonoticity has to be interpreted with considerable care.

[Table 1]

Independence was more commonly satisfied than it was violated. 58% of respondents showed an equal proportion of violations of comonotonic and noncomonotonic independence in their choices. This means under the null hypothesis that they behaved as EU maximizers. This is a percentage very similar to that reported by Weber and Kirsner (1997) with monetary outcomes. 56% of the participants in that study exhibited an equal proportion of both independence violations in their choice judgments. However, these authors only used a graphical display condition for presenting the gamble pairs.

We investigated next whether some violations of independence might be due to random inconsistencies. In fact, there are error theories that assume that respondents make mistakes when they express their preferences (Hey and Orme 1994, Harless and Camerer 1994). If violations of independence are random, then the rates of violation of independence and inconsistencies, i.e. the rate of choices reversed when the pair of gambles is replicated, will not differ systematically from each other. The difference between both rates was not significant in all cases ($p > 0.10$, t-test) except for the collapsed tree display ($p < 0.05$). Thus, as a whole, violations of independence were not systematic.

Since the different sub-hypothesis did not hold, we next compared the rate of violations of independence observed on the different experimental conditions. A significant difference between the two health contexts was found ($p < 0.01$). The direction of the discrepancy (the proportion of violations is larger for derma than for cancer) is consistent with the so-called *gambling effect* (Gafni and Torrance 1984).

According to this effect subjects exhibit a systematic dislike towards risk in the medical context. It predicts that the severity of symptoms associated with cancer leads subjects to choose the safer gamble in most pairs. Obviously, the more pronounced this tendency is, the fewer the changes in preferences. The high correlation found between proportions of independence violations and proportions of safer choices supports this explanation (ρ= -0.89, $p<.0001$). Significant differences were not found when violation rates under health states A and B were compared ($p > 0.05$, paired t-test and Wilcoxson test). However, we obtained evidence that different presentations had an influence on the rate of violations of independence ($p < 0.01$, one-way ANOVA test and Kruskall-Wallis H-test). Multiple comparisons among the different displays according to the Sheffé method revealed that the violation rate of the collapsed tree format was significantly larger than the remaining displays ($p < 0.01$). Moreover, the rate of inconsistency and the average time to complete the questionnaire were also much larger in the collapsed tree format. Thus, it seems that the complexity of the collapsed presentation affected violation and inconsistencies rates. Fennema and Wakker (1996) also found evidence on presentation effects for decision under ambiguity and monetary outcomes. However, we note that comparable results were not reported by Wakker et al. (1994).

5. Discussion

This paper provides a critical test of RDU in the domain of (risky) health outcomes. The experiment suggests that RDU, in its more general formulation, does not outperform EU. We expected that RDU would be of descriptive value in the medical context. This was the main motivation of our investigation. The same negative result for RDU was found for money by Wakker et al. (1994) and Fennema and Wakker (1996).

The support found in this experiment for the independence assumption is similar to that found by Treadwell (1998) for the preferential independence condition. These results are consistent with a "EU plus noise" model in such a way that subjects evaluate (both risky and riskless) health outcomes in an expected utility fashion but they state preferences with a random error (noise). On the contrary, Bleichrodt and Pinto (1999) found no evidence on “propagation of error” to elicit the probability weighting function in a medical decision analysis context.

However, we also found an interesting finding that qualifies the apparent rejection to the general RDU model. It seems that those respondents assigned to the display in which the commonality of the outcomes was not obvious (collapsed presentation) were indeed sensitive to rank-dependence. RDU provided descriptive improvement over EU in this display. This finding is consistent with the explanation of the Wakker et al. null result suggested by Weber and Kirsner (1997). According to this explanation, people consistently edit and cancel common outcomes in choice pairs when the framing of the decision problems is transparent enough. This editing and cancellation operation is a component of both Prospect Theory (Kahneman and Tversky 1979) and CPT (Tversky and Kahneman 1992) and empirical evidence about it was reported by Wu (1994). If subjects cancelled common outcomes in our experiment, then the four pairs of gambles within each set would become identical and preference would not vary from one to another, except for random error. Obviously, if violations were random, there would be no significant difference in the overall rate of comonotonic versus noncomonotonic independence violations. Therefore, cancellation would not allow us to test whether respondents apply the RDU rule to evaluate transparent gambles.

Heuristics of choice like editing and cancellation of common outcomes may have distorted our test of comonotonic independence. In fact, there is recent evidence on the influence of framing effects on the performance of RDU models in the medical decision making literature. Stalmeier and Bezembinder (1999) found evidence that sensitivity towards framing and probability weighting altogether can explain the discrepancy between risky and riskless utilities. Bernstein et al. (1999) found empirical support for *editing effects* in the (CPT) gain-losses health outcomes domain. Since we have found positive (but marginal) evidence for comonoticity when a non-transparent display was used, it would be worth for future research to clarify if RDU would outperform EU using other stimuli design. In this sense, Birbaum and McIntosh (1996) provide an interesting example. These authors intermixed a large number of filler pairs, in which all outcomes were different, with the experimental gamble pairs. They observed that this distribution of stimuli restrained the cancellation of common outcomes, founding positive evidence of rank-dependence. The inclusion of filler pairs and the use of non-transparent experimental gamble pairs to test RDU versus EU in the medical field is a topic for future research.

The cancellation heuristic can also explain why Bleichrodt and Pinto (2000) elicited an inverse S-shaped probability weighting function in the health domain but we could not find support for comonotonic independence in the same context. People may transform probabilities in a rank-dependent manner but the crucial test that leads to this transformation fails because subjects simplify their task canceling out the common outcome. However, it can also be true that people can transform probabilities following other theory of choice different to RDU. At this stage, the only thing that seems clear is that there seems to be a contradiction between the positive evidence in favour of

probability transformation and the negative evidence against comonotonic independence.

Appendix 1. Description of the health states.

SYMPTONS			
CANCER		DERMATITIS	
HEALTH STATE A	HEALTH STATE B	HEALTH STATE A	HEALTH STATE B
1. Slight hoarseness 2. Moderate discomfort in eating 3. No pain in breathing 4. No loss of blood	1. Mild hoarseness 2. Moderate discomfort in eating 3. Moderate pain in breathing 4. No loss of blood	1. Moderate rash 2. Slight inflammation 3. A few blisters 4. Slight skin itching	1. Extreme rash 2. Moderate inflammation 3. A few blisters 4. Slight skin itching

References

Abdellaoui, M. (2000) Parameter-Free Elicitation of Utilities and Probability Weighting Functions. *Management Science* 46 (11): 1497-1512.

Bayoumi, A. M. and Redelmeier, D. A. (2000) Decision Analysis with Cumulative Prospect Theory. *Medical Decision Making* 20: 404-412.

Berstein, L.M., Chapman, G.B. and Elstein, A. S. (1999) Framing effects in choices between multioutcome life-expectancy lotteries. *Medical Decision Making* 19 (3): 324-38.

Bleichrodt, H. (1996) Applications of utility theory in the economic evaluation of health care. *Thesis Erasmus University Rotterdam*, Rotterdam, The Netherlands.

Bleichrodt, H. (1999) Loss aversion and probability weighting in health care evaluation. *Mimeo*.

Bleichrodt, H. and Quiggin, J. (1997) Characterizing QALYs under a general rank-dependent utility model. *Journal of Risk and Uncertainty* 15(2): 151-166.

Bleichrodt, H. and Pinto, J. L. (2000) A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis. *Management Science* 46 (11): 1485-1496.

Bleichrodt, H., van Rijn, J. and Johannesson, M. (1999a) Probability weighting and utility curvature in QALY-based decision making. *Journal of Mathematical Psychology* 43: 238-60.

Bleichrodt, H., Pinto J. L. and Wakker, P. P. (1999b) Using Descriptive Findings of Prospect Theory to Improve Prescriptive Applications of Expected Utility (*working paper*). Department of Medical Decision Making, University Medical Center, Leiden, The Netherlands.

Birnbaum, M. H. and McIntosh, W. R. (1996) Violations of branch independence in

choices between gambles. *Organizational behavior and human decision processes* 1(67): 91-110.

Camerer, C. F. (1989) An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty* 2: 61-104.

Camerer, C. F. (1992) Recent tests of generalization of expected utility theory. In: Edwards, W. (ed.) *Utility theories: measurements and application.* Boston: Kluwer Academic Publishers.

Camerer, C.F. (1995) *Individual decision making*. In: Kagel, J. y Roth, A. (eds.) *The handbook of experimental economics.* Princeton UP, USA.

Currim, I. S. and Sarin, R. K. (1989) Prospect Versus Utility. *Management Science* 35: 22-41.

Fennema, H. and Wakker, P. (1996) A test of rank-dependent utility in the context of ambiguity. *Journal of Risk and Uncertainty* 13: 19-35.

Gafni, A. and Torrance, G. W. (1984) Risk attitude and time preference in health. *Management Science* 30(4): 440-451.

Harless, D. and Camerer, C. F. (1994) The Predictive Utility of Generalized Expected Utility Theories. *Econometrica* 62: 1251-1289.

Hey, J. and Orme, C. (1994) Investigating Generalizing of Expected Utility Theory Using Experimental Data. *Econometrica* 62(6): 1291-1326.

Kahneman, D. and A. Tversky (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47(2): 263-291.

Lattimore, P. M., Baker, J. R. And Witte, A. D. (1992) The Influence of Probability on Risky Choice. *Journal of Economic Behavior and Organization* 17: 377-400.

Prelec, D. (1998) The probability weighting function. *Econometrica* 66 (3): 497-527.

Preston, M. C. and Baratta, P. (1948) An experimental study of the auction value of an

uncertain outcome. *American Journal of Psychology* 61: 183-193.

Quiggin, J. (1982) A theory of Anticipated Utility. *Journal of Economic Behavior and Organization* 3(4): 323-343,

Stalmeier, P.F. and Bezembinder, T. G. (1999) The discrepancy between risky and riskless utilities: a matter of framing?. *Medical Decision Making* 19(4): 435-47.

Treadwell, J. R. (1998) Test of preferential independence in the QALY model. *Medical Decision Making* 18(4): 418-428.

Tversky, A. and Kahneman, D. (1981) The Framing of Decisions and the Psychology of Choice. *Science* 211(30): 453-8.

Tversky, A. and Kahneman, D. (1986) Rational choice and the framing of decisions. *Journal of Business* 59 (4): S251-S277.

Tversky, A. and Kahneman, D. (1992) Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5: 297-342.

Wakker, P. (1996) The sure-thing principle and the comonotonic sure-thing principle: an axiomatic analysis. *Journal of Mathematical Economics* 25: 123-227.

Wakker, P., Erev, I. and Weber, E. U. (1994) Comonotonic independence: the critical test between classical and rank-dependent utility theories. *Journal of Risk and Uncertainty* 9: 195-230.

Wakker, P. and Stiggelbout, A. (1995) Explaining distorsions in utility elicitation through the rank-dependent model for risky choices. *Medical Decision Making* 15: 180-186.

Weber, E.U. and Kirsner, B. (1997) Reasons for rank-dependent utility evaluation. *Journal of Risk and Uncertainty* 14: 41-61.

Wu, G. (1994) An empirical test of ordinal independence. *Journal of Risk and Uncertainty* 9: 39-60.

Wu, G. and González, R. (1996) Curvature of the probability weighting function. *Management Science* 42 (12): 1676-1690.

Wu, G. and González, R. (1998) Common consequence conditions in decision making under risk. *Journal of Risk and Uncertainty* 16: 115-139.

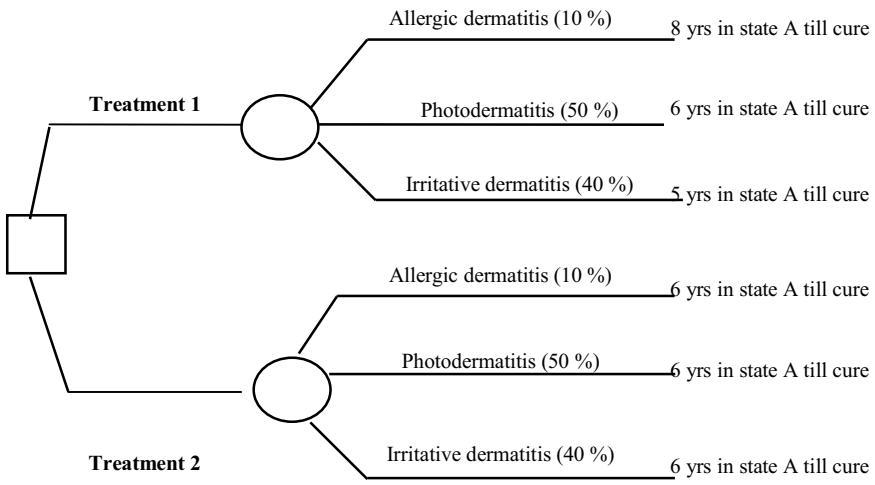
Tables and figures

Figure 1. Display formats

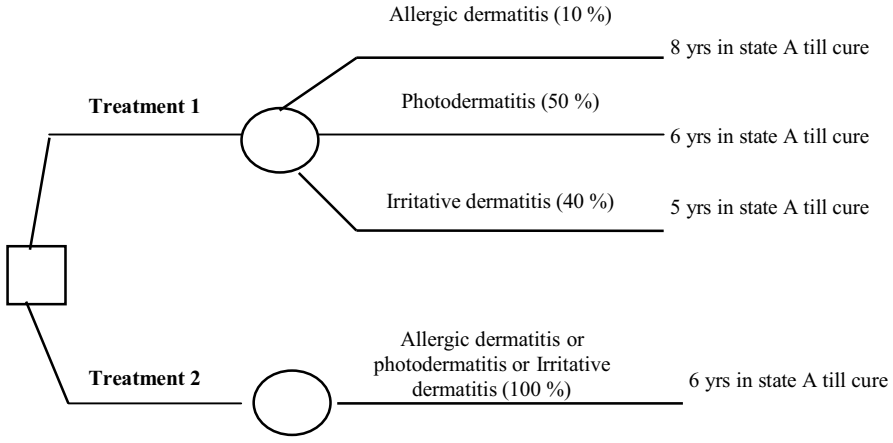
1.A Matrix

	Allergic Dermatitis	Photodermatitis	Irritative Dermatitis	
	10 %	50 %	40 %	
Treatment 1	8	6	5	Yrs in state A till cure
Treatment 2	6	6	6	Yrs in state A till cure

1.B Non-collapsed decision-tree



1.C Collapsed decision-tree



1.D Graphic

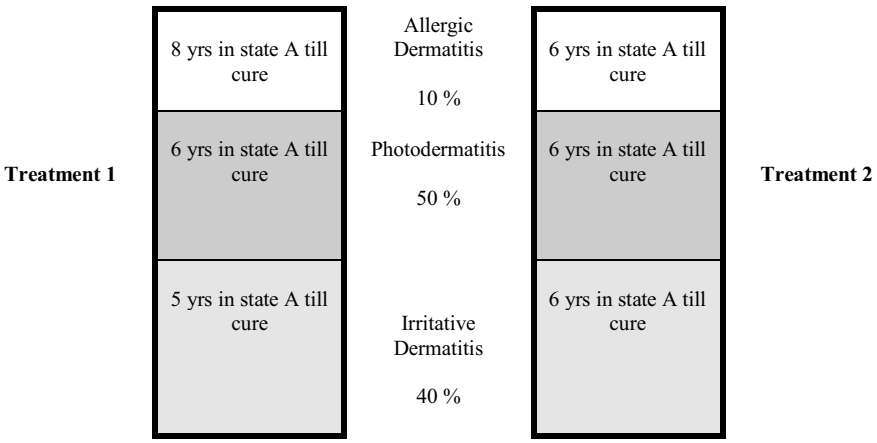

Treatment 1
8 yrs in state A till cure
6 yrs in state A till cure
5 yrs in state A till cure
Allergic Dermatitis
10 %
Photodermatitis
50 %
Irritative Dermatitis
40 %
6 yrs in state A till cure
6 yrs in state A till cure
6 yrs in state A till cure
Treatment 2

Figure 2. Sets of gambles relevant for testing comonotonic and noncomonotonic independence (Health state A)

Set 1

Event 1= irritative contact dermatitis /spread epidermoid cancer
Event 2= allergic contact dermatitis /non-spread epidermoid cancer
Event 3 = photodermatitis/esophagic adenocarcinoma

	Event 1	Event 2	Event 3
Risky gamble R	CO	7'5	4
Sure gamble S	CO	7	5

CO (common outcome)=10 → 8 → 6 → 4 (yrs in state A)
P(Event 1)= 0'7; P(E2)= 0'2; P(E3)= 0'1. P denotes probability.
EV(R)=EV(S). EV denotes expected value.

Set 2

Event 1= allergic contact dermatitis/ non-spread epidermoid cancer
Event 2= irritative contact dermatitis / spread epidermoid cancer
Event 3= photodermatitis/ esophagic adenocarcinoma

	Event 1	Event 2	Event 3
Risky gamble R	CO	9	1'5
Sure gamble S	CO	8	3'5

CO (common outcome)=10 → 7 → 4 → 1 (yrs in state A).
P(Event 1)= 0'4; P(E2)= 0'4; P(E3)= 0'2. P denotes probability.
EV(R)=EV(S). EV denotes expected value.

Set 3

Event 1= photodermatitis/ esophagic adenocarcinoma
Event 2= allergic contact dermatitis/ non-spread epidermoid cancer
Event 3= irritative contact dermatitis / spread epidermoid cancer

	Event 1	Event 2	Event 3
Risky gamble R	CO	8	5
Sure gamble S	CO	6	6

CO(common outcome)= 8 → 6 → 4 → 2 (yrs in state A).
P(Event 1)= 0'5; P(E2)= 0'1; P(E3)= 0'4. P denotes probability.
EV(R)=EV(S) – 0'2 yrs is state A. EV denotes expected value.

Set 4

Event 1= irritative contact dermatitis / spread epidermoid cancer
Event 2= allergic contact dermatitis/ non-spread epidermoid cancer
Event 3= photodermatitis/ esophagic adenocarcinoma

	Event 1	Event 2	Event 3
Risky gamble R	CO	4'5	2'5
Sure gamble S	CO	4	3

CO(common outcome)= 5 → 3.5 → 2 → 0.5 (yrs in state A).
P(Event 1)= 0'55; P(E2)= 0'15; P(E3)= 0'3. P denotes probability.
EV(R)=EV(S) – 0'075 yrs in state A. EV denotes expected value.

Table 1. Proportion of observed violations of comonotonic independence (Pci obsd.) versus proportion of violations of comonotonic independence predicted (Pci pred.) by the null hypothesis[1]

	Different experimental conditions	Violations Comono-tonic ind. (Vci)	Violations Noncomo-notonic ind. (Vnci)	Violations EU inde-Pendence (Vind=Vci+ Vnci)	Pci obsd. (Vci/Vind)	Difference with Pci pred. (1/3-Vci/Vind)
Sub-Hypothesis 1	Cancer	14	30	44	0,32	0,01
	Derma	24	50	74	0,32	0,01
Sub-Hypothesis 2	Health state A	21	39	60	0,35	-0,02
	Health state B	17	41	58	0,29	0,04
Sub-Hypothesis 3	Matrix	7	13	20	0,35	-0,02
	Graphic	8	12	20	0,40	-0,07
	Tree	11	19	30	0,37	-0,04
	Collapsed tree	11	37	48	0,23	0,10§

[1] H_0 predicts that the proportion of violations of comonotonic independence = 1/3.
[2] § denotes a significant difference at the 5% level.