Journal of Experimental Psychology: General
1988, Vol. 117, No. 1, 3–20

0096-3445/88/$00.75

A Multiplicative Model of the Utility of Survival Duration and Health Quality

John M. Miyamoto
University of Washington

Stephen A. Eraker
Department of Medicine
Veterans Administration Medical Center
Salt Lake City, Utah

SUMMARY

Survival duration and health quality are fundamentally important aspects of health. A utility model for survival duration and health quality is a model of the subjective value of these attributes. We investigate the hypothesis that the utility (subjective value) of survival duration and health quality is determined by a multiplicative model. According to this model, there are separate subjective scales for the utility of survival duration and health quality. If $F(Y)$ equals the utility of surviving Y years, and $G(Q)$ equals the utility of living in health state Q, then the multiplicative model proposes that $F(Y)G(Q)$ equals the utility of surviving Y years in health state Q. This model provides a simple explanation for several intuitively compelling relationships. First, the distinction between better-than-death and worse-than-death health states corresponds to the assignment of positive or negative utilities to different health states. Second, a zero duration of survival removes any reason to prefer one health state over any other, just as multiplying the utility of health quality by zero eliminates differences between the utilities of different health states. Third, the subjective difference between Y years in pain and Y years free from pain increases as Y increases as if the difference in utility between pain and no pain were being multiplied by the utility of surviving Y years. A critical prediction of the multiplicative model is the hypothesis that preferences between gambles for health outcomes satisfy a property called utility independence. After defining this property and explaining how it can be tested in behavioral data, we report an experimental test of whether the health preferences of medical patients satisfy utility independence. Individual analyses revealed that most subjects satisfy utility independence, thereby supporting the multiplicative utility model. Some subjects appear to violate a fundamental assumption of utility theory: They appear to violate the assumption that a single utility scale represents both the ordinal preference relations between certain outcomes and the subjective averaging that underlies the utility of gambles. The violation is inferred from an inconsistency between preferences for multiattribute outcomes when they are viewed as certain outcomes and when they are viewed as the outcomes of gambles.

When choosing a medical therapy, survival duration and health quality are two of the most important attributes of the outcome of a decision. To make an optimal choice when duration and quality of survival are at issue, it is necessary to evaluate the subjective value of the combinations of duration and quality that could result from a choice. The theory of expected utility is a normative theory of decision making in which a measure of subjective value, called *utility*, is defined (Luce & Raiffa, 1957; Von Neumann & Morgenstern, 1947). When applied to the problem of therapy selection, expected utility theory describes a method for determining the choice that maximizes the probabilistic expectation of the utility to the patient (Weinstein et al., 1980). Individuals differ in their willingness to undergo risks to achieve potential improvements in health, or in the trade-offs that they make between the duration and quality of survival (McNeil, Weichselbaum & Pauker, 1978, 1981; Miyamoto & Eraker, 1985). The optimal choice for one patient may not be optimal for another; one purpose of the utility analysis of health outcomes is to represent individual differences in values for health outcomes (McNeil & Pauker, 1982; Pauker & McNeil, 1981).

In the present essay we will not attempt to discuss the complete decision analysis of therapeutic choices; rather, we focus on the particular problem of testing a utility model for

---

We gratefully acknowledge the advice and criticism of Clyde Coombs, Richard Cornell, David H. Krantz and J. Frank Yates. John Clavadetscher, Bob Frick, Aura Hanna, Earl Hunt and Tom Nelson provided useful comments on a draft of this essay. We would like to thank Susan Baker for able assistance in carrying out patient interviews. A preliminary version of this research was presented in a doctoral dissertation by John M. Miyamoto.

Correspondence concerning this article should be addressed to John M. Miyamoto, Department of Psychology NI-25, University of Washington, Seattle, Washington 98195.

survival duration and health quality. The utility of survival duration and health quality is clearly of practical importance. We will show that preferences for survival duration and health quality are also of theoretical interest.

The primary hypothesis investigated here is the hypothesis that the utility of survival duration and health quality is described by a multiplicative model. To state this hypothesis precisely, it will help to introduce some formal notation. Let (Y, Q) denote a health outcome consisting of a survival lasting Y years in health state Q, followed by death at the end of the Yth year. For example (10 years, moderate pain), denotes a survival lasting 10 years in moderate pain, followed by death at the end of the 10th year. For purposes of the present study, we assume that the health state Q is constant during survival. Although the health outcomes studied here are only a subset of naturally occurring health outcomes (outcomes in which health quality changes during survival are omitted), preferences for these simplified outcomes exhibit phenomena that are important in their own right. Furthermore, the utility analysis of survival in constant health states can be a useful approximation in medical decision analyses (Miyamoto & Eraker, 1985; Pliskin, Shepard, & Weinstein, 1980; Weinstein et al., 1980; Weinstein & Stason, 1982).

Let $U(Y, Q)$ denote the utility of the (Y, Q) outcome. A multiplicative utility health postulates the existence of subjective scales, F and G, that apply to Y and Q, respectively, such that

$$U(Y, Q) = F(Y)G(Q) \tag{1}$$

for every Y and Q. In other words, the multiplicative utility model proposes that there exist subjective scales F and G such that $F(Y)$ is the utility of living Y years, $G(Q)$ is the utility of health state Q, and the utility of living Y years in health state Q is the product $F(Y)G(Q)$.

The quantity $G(Q)$ can be interpreted as a discounting factor for the utility of surviving in health state Q (Miyamoto & Eraker, 1985). For example, suppose that G(normal health) = 1.00, G(mediocre health) = .75 and G(poor health) = .50; then,

$$U(Y, \text{normal health}) = F(Y)(1.00),$$
$$U(Y, \text{mediocre health}) = F(Y)(.75), \text{ and}$$
$$U(Y, \text{poor health}) = F(Y)(.50).$$

In other words, the utility of Y years in mediocre health is 75% of the utility of Y years in normal health, and the utility of Y years in poor health is 50% of the utility of Y years in normal health. The multiplicative utility model conceptualizes the utility of (Y, Q) as determined by the utility of living Y years, $F(Y)$, which is discounted by the factor $G(Q)$, the relative worth of the health state Q. In this interpretation, it is natural to think of $G(Q)$ as having a value of 1.0 when Q is normal health (however one wants to define it). The utility of any other health state is a percentage of the utility of normal health.

There are several intuitively compelling reasons for suspecting that the utility of survival duration and health quality is determined by a multiplicative model. First, longer survival is usually preferred to shorter survival, but one may prefer shorter survival to longer survival if the health state is extremely undesirable. In other words, some health states are regarded as worse than death. A multiplicative utility model provides a simple way to account for this relation. Suppose that $U(Y, Q) = F(Y)G(Q)$ and that $F(Y)$, the utility of Y years, grows larger as Y grows larger. If $G(Q) > 0$, the combined utility $F(Y)G(Q)$ is an increasing function of duration, but if $G(Q) < 0$, then the combined utility, $F(Y)G(Q)$, is a decreasing function of duration. Multiplying $F(Y)$ by a negative number inverts the ordering produced by variation in Y. The relationship that longer survivals are preferred in some health states and shorter survivals are preferred in others is represented in the multiplicative utility model by the sign of $G(Q)$.

Second, consider the effect of survival duration on the utility of health states. At positive durations of survival, one has preferences between different health states. For example, at 3 years duration, "no pain" is preferred to "slight pain" which is preferred to "great pain." Symbolically, the relation is

$$U(3 \text{ years, no pain}) > U(3 \text{ years, slight pain}) > U(3 \text{ years, great pain}).$$

At 0 duration, however, one is indifferent between the three health states. In other words,

$$U(0 \text{ years, no pain}) = U(0 \text{ years, slight pain}) = U(0 \text{ years, great pain}).$$

The multiplicative utility model also provides a simple account for this relationship. Assume that $F(Y) > 0$ when $Y > 0$, but $F(0) = 0$. Then, $U(Y, Q) = F(Y)G(Q)$ will vary as a function of quality when $Y > 0$, but when $Y = 0$, $U(0, Q) = F(0)G(Q) = 0 \cdot G(Q) = 0$ is a constant function of quality. Multiplying $G(Q)$ by zero nullifies the ordering produced by variation in Q, just as a zero duration removes any reason to prefer one health quality over another.

A third reason for postulating a multiplicative utility model is that subjective differences between health states are increased as the duration of survival increases. For example, the subjective difference between 10 years in excellent health and 10 years in mediocre health is greater than the subjective difference between 1 year in excellent health and 1 year in mediocre health. To express this symbolically, let EH denote excellent health and let MH denote mediocre health. It follows that

$$U(10 \text{ years}, EH) - U(10 \text{ years}, MH) > U(1 \text{ year}, EH) - U(1 \text{ year}, MH). \tag{2}$$

In general, the utility difference between Y years in excellent health and Y years in mediocre health is small when Y is short, and it grows larger as Y increases.

To see how the multiplicative utility predicts this relation, note that the $F(Y)$ grows larger as Y grows larger. Hence, $F(10 \text{ years}) > F(1 \text{ year})$, and thus

$$F(10 \text{ years})[G(EH) - G(MH)] > F(1 \text{ year})[G(EH) - G(MH)]. \tag{3}$$

But the multiplicative utility model implies that

$$\begin{aligned} &\mathrm{F}(10 \text{ years})[G(EH) - G(MH)] \\ &\quad = F(10 \text{ years})G(EH) - F(10 \text{ years})G(MH) \qquad (4) \\ &\quad = U(10 \text{ years}, EH) - U(10 \text{ years}, MH), \end{aligned}$$

and

$$\begin{aligned} &\mathrm{F}(1 \text{ year})[G(EH) - G(MH)] \\ &\quad = F(1 \text{ year})G(EH) - F(1 \text{ year})G(MH) \qquad (5) \\ &\quad = U(1 \text{ year}, EH) - U(1 \text{ year}, MH). \end{aligned}$$

Substituting Equations 4 and 5 into Equation 3 yields $U(10 \text{ years}, EH) - U(10 \text{ years}, MH) > U(1 \text{ year}, EH) > U(1 \text{ year}, MH)$, in conformity with the intuition expressed in Equation 2. The multiplicative utility model predicts that utility differences between health states appear greater when the associated survival duration is longer, because $U(Y, Q_1) - U(Y, Q_2) = F(Y)[G(Q_1) - G(Q_2)]$ and $F(Y)$ grows larger as Y grows larger.

These arguments demonstrate the plausibility of a multiplicative utility model. The main goal of this essay is to formulate a rigorous test of the multiplicative utility model, and to carry out the test experimentally. Although the applied significance of a health utility model is part of the motivation for the present study, we will not focus on issues of medical decision analysis. The heart of our study is the formulation of five postulates that are sufficient for the validity of the multiplicative utility model in the sense that if the five postulates are empirically valid, the model is necessarily true. Each postulate is a hypothesis concerning preference behavior or the mental representations that underlie preference behavior. The advantage of formalizing the model as a system of postulates is that it directs attention to critical behavioral tests of the model (Coombs, 1983; Krantz, 1974). Furthermore, if violations of predictions are found, it is possible to use the formal analysis to determine which assumptions are suspect in the light of failed predictions.

Our analysis is developed within the framework of conjoint measurement theory (Krantz, Luce, Suppes, & Tversky, 1971); it also incorporates more recent work in utility theory (Kahneman & Tversky, 1979; Keeney & Raiffa, 1976; Luce & Narens, 1985; Miyamoto, in press). By necessity, the theory and issues that we present require a mathematical treatment, but the discussion does not assume an advanced understanding of conjoint measurement theory.

The structure of the essay is as follows. We first present three postulates that describe preference behavior in general, rather than the specific nature of health preferences. These postulates are widely shared by theories of preference under risk (preference between gambles). They are not the focus of our analysis, but rather serve as background to utility assumptions that are specific to survival duration and health quality. The fourth postulate is the assumption that health qualities are equally valued when survival duration is zero. This assumption has already been discussed. Finally, we describe a property of preference judgment, called *the utility independence of survival duration from health quality*, which is predicted by the multiplicative utility model. Our fifth postulate is the hypothesis that preferences satisfy utility independence.

The utility independence property is a critical prediction of the multiplicative utility model. An experimental test of utility independence will be reported in which medical patients were asked to judge preferences for gambles for health outcomes. The experimental results will be discussed with respect to the specific question, whether the utility of survival duration and health quality is multiplicative, and also with respect to the general problem of constructing and testing multiattribute utility models.

Background Assumptions From Utility Theory

Our first three postulates are weak assumptions that are implied by a variety of theories of preference under risk (preference between gambles). To explain the rationale for these postulates, we must digress briefly into theories of utility and preference under risk.

In expected utility theory, the utility of outcomes is inferred from preferences between hypothetical gambles (Luce & Raiffa, 1957; Von Neumann & Morgenstern, 1947). If preferences between gambles satisfy the fundamental assumptions (axioms) of expected utility theory, then there exists a utility scale such that the utility of a gamble is a weighted average of the utilities of its separate outcomes (Luce & Raiffa, 1957). The details of expected utility theory are not important here. What is relevant, however, is that empirical predictions of the theory have since been shown to be false (Kahneman & Tversky, 1979; MacCrimmon & Larsson, 1979; Tversky, 1975), and a number of alternative theories have been proposed to overcome deficiencies of expected utility theory (Kahneman & Tversky, 1979; Karmarkar, 1978; Luce & Narens, 1985). The alternative theories all have much in common with expected utility theory, but they deviate in ways that attempt to avoid the disconfirmed predictions of expected utility theory.

The theoretical assumptions on which we develop the present investigations can be thought of as "generic" assumptions in the following sense. Rather than selecting one of the revisionist theories as the framework for our investigations, we base our work on three assumptions (Postulates 1–3) that are implied by all of these theories. In particular, these assumptions are implied by Karmarkar's (1978) subjectively weighted utility model, Luce and Narens's (1985) dual bilinear model, and Savage's (1954) subjective expected utility theory, as well as by expected utility theory. Furthermore, the assumptions are implied by Kahneman and Tversky's prospect theory when certain, easily satisfied conditions hold.[1] Postulates 1–3 are generic assumptions in the sense that they are weak assumptions that are implied by many theories without asserting anything that is peculiar to only one theory. By basing our work on weak assumptions, it is possible to interpret our results from the standpoint of all of these theories without being committed to any one of them.

---

[1] The relevant conditions are spelled out in the discussion of prospect theory in the Results section.

Our assumptions are formulated in terms of even-chance gambles (two-outcome gambles with a 50% chance of receiving either outcome). We restrict the theory to even-chance gambles because we are attempting to study the utility of outcomes, and not the subjective representation of probability, and it is simpler to study gambles with a single, fixed probability. Let $(A, .5, B)$ denote an even-chance gamble for outcomes A and B, and let $\geq_p$ stand for an individual's preference ordering over such gambles. In other words, $(A, .5, B) \geq_p (C, .5, D)$ indicates that the individual regards $(A, .5, B)$ to be at least as desirable or more desirable than $(C, .5, D)$. It will simplify the discussion to assume that preference judgments are free from random error. The problem of testing preference assumptions in the context of response variation will be addressed later. We can now state Postulates 1–3 of our analysis.

First, we assume there exists a scale U that represents ordinal preference relations in the following sense.

Postulate 1. $(A, .5, B) \geq_p (C, .5, D)$ iff $U(A, .5, B) \geq U(C, .5, D)$. Note that a gamble of the form $(X, .5, X)$ is a gamble in which one is certain to receive X. Hence, $A = (A, .5, A)$, $B = (B, .5, B)$, and the following is a special case of Postulate 1.

Postulate 1′. $A \geq_p B$ iff $U(A) \geq U(B)$.

Thus, Postulate 1 implies that the utility scale U represents both the preference order for risky options (gambles) and the preference order for riskless options (outcomes).

Second, we assume that the utility of a gamble is a weighted average of the utilities of its outcomes. When evaluating the utility of $(A, .5, B)$, the individual assigns a subjective weight $W(.5)$ to the .5 probability $(0 < W[.5] < 1)$ and combines the utilities of A and B by the following rule:

$$U(A, .5, B) = W(.5)U(A) + [1 - W(.5)]U(B). \quad (6)$$

Because the present theory is restricted to even-chance gambles, we will adopt a simplified notation in which $s = W(.5)$ and $t = 1 - W(.5)$. In this notation, Equation 6 is equivalent to the following.

Postulate 2. There exist constants s and t such that $0 < s < 1$, $t = 1 - s$, and

$$U(A, .5, B) = sU(A) + tU(B). \quad (7)$$

According to expected utility theory, the weights, s and t, must satisfy $s = t = .5$, but this restriction is not imposed here.

In order to make Postulate 2 consistent with the dual bilinear model, we will assume that it only applies to gambles of the form $(A, .5, B)$ such that $A \geq_p B$. For example, we assume that Postulate 2 applies to the gamble (10 years, .5, 6 years), but not to the gamble (6 years, .5, 10 years). The reason for restricting Postulate 2 in this way is that the dual bilinear model allows for the possibility that the weights, s and t, could have different values depending on whether $A \geq_p B$ or $A \leq_p B$ (Luce & Narens, 1985; Narens & Luce, 1986). For example, it might be the case that $U(A, .5, B) = .3U(A) + .7U(B)$ when $A \geq_p B$, but $U(A, .5, B) = .6U(A) + .4U(B)$ when $A \leq_p B$. Even with the probability held fixed at .5, different subjective weights can apply to an outcome depending on whether it is the more desirable or less desirable outcome of the gamble. Thus, the dual bilinear model does not predict that Postulate 2 will be satisfied by *all* even-chance gambles, but it does predict that it will be satisfied if we restrict the postulate only to gambles such that $A \geq_p B$, or only to gambles such that $A \leq_p B$. We have chosen to adopt the former restriction.

The restriction of Postulate 2 to gambles $(A, .5, B)$ such that $A \geq_p B$ also makes possible an interpretation of the postulate from the standpoint of prospect theory. Because this interpretation is complex, we will defer its discussion to the Results section, where prospect theory will be discussed in a unified manner.

Our third assumption is a continuity assumption.

Postulate 3. Either the utility scale U is continuous, or else the number of points at which it is discontinuous is finite.

Postulate 3 is always reasonable in any application of utility theory. Functions that violate Postulate 3 simply do not occur as subjective representations. The importance of Postulate 3 is that in combination with Postulates 1 and 2, it implies that the utility scale is an interval scale (Miyamoto, 1985, in press).[2] The interval scale property plays a central role in the mathematical derivation of the multiplicative utility model (Keeney & Raiffa, 1976; Miyamoto, 1983, 1985, in press), but this derivation will not be discussed here.

Postulates 1–3 are not necessary for the multiplicative utility model, but in combination with Postulates 4 and 5 described later, they are sufficient for the model. Postulates 1 and 2 are implied by all the main theories of preference under risk, in particular, by the theories listed at the beginning of this section. Postulate 3 is not strictly implied by these theories, but it is extremely plausible from every theoretical standpoint. In subsequent discussions, we will generally say that Postulates 1–3 are "implied" by theories of preference, without stating more precisely that Postulates 1 and 2 are literally implied by these theories, whereas Postulate 3 is only extremely plausible. The reason we emphasize these logical relations between Postulates 1–3 and theories of preference under risk is that the theoretical and empirical results that we report can be interpreted from the standpoint of any theory that implies these postulates. By basing our working on a weak set of assumptions (Postulates 1–3), we broaden the class of theories relative to which results can be interpreted.

Zero Duration Nullifies Preference Over Health Quality

Our fourth postulate is the assumption that all health qualities are equally valued if the duration of survival is zero. Because we have already discussed this postulate, it only remains to state it formally. Let $\sim_p$ stand for equality in preference.

Postulate 4. All health qualities are equally valued when survival duration is zero. Stated formally, for any health

---

[2] In a standard conjoint measurement formulation, Postulates 1–3 would not be taken as basic assumptions. Rather, qualitative preference assumptions would be formulated from which Postulates 1–3 can be derived. Such assumptions have been formulated in Miyamoto (1985, in press), but their statement has been omitted here for the sake of simplicity.

qualities Q_1 and Q_2,

$$(0, Q_1) \sim_{p} (0, Q_2). \qquad (8)$$

Postulate 4 is a necessary condition for a multiplicative utility model if we assume that $F(0) = 0$ (the utility of zero duration is zero). We previously noted that the multiplicative utility model is also suggested by the existence of worse-than-death health states and by the fact that increased survival duration magnifies differences between health qualities. These assumptions have not been chosen as formal postulates because they are not needed in the derivation of the multiplicative utility model. It would be reasonable to adopt these assumptions as formal postulates but they would be redundant in the mathematical analysis.

Certainty Matching and Utility Independence

In order to state Postulate 5, we need to describe a kind of preference judgment called certainty matching. This judgment is also the primary dependent variable in the experiment reported later.

Let g be an even-chance gamble in which both survival outcomes are accompanied by the health quality Q, that is,

$$g = [(Y_1, Q), .5, (Y_2, Q)].$$

Now consider a task where a subject is asked to state a duration Y_3 such that the subject is indifferent between receiving (Y_3, Q) with certainty, and playing g. For example, suppose Q stands for "moderate pain," and Y_1 and Y_2 are 5 and 20 years, respectively. Y_3 is the response to the question, "How many years in moderate pain would be equal in value to an even-chance gamble between 5 years and 20 years in moderate pain?" Let

$$(Y_3, Q) \sim_{p} [(Y_1, Q), .5, (Y_2, Q)] \qquad (9)$$

signify that the subject has stated that (Y_3, Q) is equal in value to the gamble. The outcome (Y_3, Q) in Equation 9 is a judgment of a match, namely, a judgment of the outcome that matches the gamble in subjective value. The duration Y_3 is the dependent variable in this judgment, and Y_1, Y_2, and Q are independent variables.

There is a slightly different notation for Equation 9 that is more perspicuous. Because Q is the health state accompaning every survival outcome in Equation 9, we can write the equation in the form

$$Y_3 \sim_{p} (Y_1, .5, Y_2) \text{ in health state } Q. \qquad (10)$$

When the relation in Equation 10 holds, we will say that Y_3 is the *certainty match* of the gamble $(Y_1, .5, Y_2)$ in health state Q.

Although Equations 9 and 10 are simply alternative notations, Equation 10 makes salient the fundamental issue addressed by the utility independence property, namely, does the value of the certainty match Y_3 depend on the particular quality Q, or is it independent of it? Utility independence is the hypothesis that the certainty match is independent of health quality. Because of its importance to this work, we will state the property as a formal definition.

Definition 1. Survival duration is said to be utility independent of health quality provided that for any $(Y_1, .5, Y_2)$, the duration Y_3 that satisfies

$$Y_3 \sim_{p} (Y_1, .5, Y_2) \text{ in health state } Q.$$

is the same for any choice of Q.

Our fifth postulate is simply that survival duration is utility independent from health quality.

Postulate 5. Survival duration is utility independent from health quality in the sense of Definition 1.

Leaving aside the issue of response variability, the basic idea in testing utility independence would be to task an individual to produce certainty matches for a set of gambles for survival duration, where the health quality associated with the survival outcomes is fixed first at level Q_1, then at level Q_2, then at level Q_3, and so on. Utility independence asserts that certainty matches remain the same for different choices of health quality.

If Postulates 1 and 2 are valid, then the multiplicative utility model implies that survival duration is utility independent of health quality. (The proof is presented as Proof 1 of the Appendix). Because we are assuming Postulates 1 and 2 on the basis of prior theory, utility independence of survival duration can be regarded as a critical prediction of the multiplicative utility model. This prediction will be the central issue of the experiment reported later.

We should mention several qualifications with respect to the formulation of utility independence given here. First, the roles of survival duration and health quality could be interchanged in this formulation to yield a definition of the utility independence of health quality from survival duration. In other words, a certainty match on the quality dimension at a fixed duration Y is defined to be a quality Q_3 such that

$$Y, Q_3) \sim_{p} [(Y, Q_1), .5, (Y, Q_2)]. \qquad (11)$$

Health quality is utility independent from survival duration provided that the Q_3 satisfying Equation 11 depends on the choice of Q_1 and Q_2, but not on the choice of Y. An analogous argument to Proof 1 of the Appendix shows that the utility independence of health quality from survival duration is also predicted by the multiplicative utility model.

We have not adopted the utility independence of health quality from survival duration as an assumption of our formalization because it is not needed to derive the multiplicative utility model. Furthermore, it is harder to test than the utility independence of survival duration from health quality. To test the utility independence of health quality from survival duration, one would have to describe a larger number of health qualities in sufficient detail for their relative worths to be assessed. One would then pair these health qualities with a few survival durations, and test whether certainty matches on the quality dimension are independent of survival duration. The difficulty lies in describing a large number of health qualities within the practical limits of an interview. Each quality requires time to explain in detail. In contrast, durations of survival are easy to explain and hence the elicitation of certainty matches on the duration dimension poses fewer practical difficulties. Because it is practically difficult to test the utility independence of health quality from survival duration, we have developed a set of postulates for the multiplicative utility model that omits this prediction of the model.

We should also mention that utility independence as developed here is a special case of more general forms of utility independence (Keeney & Raiffa, 1976). Essentially, our definition is restricted to two attributes and the certainty matching judgment, but more general definitions are available for larger numbers of attributes and a broader range of preference judgments. Only Definition 1 is given here because it suffices for the present theoretical development, and it succinctly captures the main concept of utility independence. Henceforth, for the sake of brevity, we will refer to the condition stated in Definition 1 as *the* utility independence property, without repeating the qualification that it is only a particular form of utility independence.

Sufficiency of Postulates 1–5

Miyamoto (1985) proved that Postulates 1–5 are sufficient to establish the validity of the multiplicative utility model. We will state this result as a formal theorem, but before stating the theorem, we should explain a limitation on the survival durations to which the theorem applies. Survival durations are assumed to be drawn from an interval between zero and M, where M is the longest duration that is reasonable or relevant to the individuals whose preferences are being investigated. For example, if the individuals are approximately 40 years old, only survivals less than 60 years would be of primary interest. Hence, M could be set to 60 years. If the primary concern were the utility of survival between 40 and 70 years, then M could be set to 30 years. Essentially, M sets boundary beyond which the utility model is no longer claimed to apply (e.g., it does not apply to survivals in the range 200 to 1,000 years).

Theorem 1. Let $\Gamma = \{Q_1, Q_2, Q_3, \ldots\}$ be a finite or infinite set of health states. Let M be a duration that is chosen to be the longest that is relevant to the utility problem under investigation. If Postulates 1–5 hold for every health state in Γ and every duration between 0 and M, then there exist functions F and G such that $U(Y, Q) = F(Y)G(Q)$ for every Q in Γ and every Y between 0 and M.[3]

Pliskin et al. (1980) were the first to propose a multiplicative utility model for survival duration and health quality. Working in the expected utility framework, they postulated that survival duration is utility independent from health quality, and also that health quality is utility independent from survival duration. Their formalization also included the assumption that a condition called marginality or additive independence is violated. We will not attempt to define this condition here. They derived the multiplicative utility model from these assumptions and the background assumptions of the expected utility theory. Pliskin et al. did not test whether the utility independence assumptions were empirically satisfied.

Miyamoto (1985) pointed out that the existence of worse-than-death health states and the fact that a zero duration nullifies preference over health quality provide natural motivations for a multiplicative utility representation. Furthermore, he showed that Postulate 4, that zero duration nullifies preference over quality, eliminates the need for the assumptions that health quality is utility independent of survival duration and that the marginality condition is violated. Miyamoto and Eraker (1985), Pliskin et al. (1980), and Weinstein et al. (1980) discussed applications of the multiplicative utility model to the utility analysis of medical therapy selection. Miyamoto (1985, in press) pointed out that the formalization of the multiplicative utility model could be developed under weak background assumptions that are similar to Postulates 1–3. The advantage of this approach is that the multiplicative utility model and other multiattribute models can be formalized and tested under assumptions that are compatible with a broader range of theories than expected utility theory.

Testing the Formalization

To test Postulates 1 and 2, one would first have to formulate observable properties of preference behavior from which these postulates could be derived. Although such properties have been formulated (Miyamoto, 1985, in press), they will not be stated or tested here. If we are in error in assuming Postulates 1 and 2 on the basis of prior theory, then the theories that imply them will have to be discarded or revised, and further research into utility will be conducted in a greatly altered theoretical framework. Without wishing to be confusing, we should nevertheless mention that the experimental results reported later do provide some intriguing evidence against Postulates 1 and 2. Although our experiment was not designed to test these postulates, some of the results are difficult to account for if Postulates 1 and 2 are true. Later, we will sketch how violations of Postulates 1 and 2 might arise. It will be helpful to keep in mind, however, that our experimental work was designed under the assumption that Postulates 1 and 2 are valid.

Postulates 3 and 4 will not be tested here. The assumption that the utility function is either continuous, or else it is discontinuous at finitely many points (Postulate 3) is obviously satisfied by any psychologically plausible utility function. The assumption that all health qualities are equally valued if survival duration is zero (Postulate 4) is introspectively obvious. Thus, utility independence (Postulate 5) is the primary candidate for empirical test. It is a critical test of the multiplicative model because, under the assumption of Postulates 1–4, utility independence is both necessary and sufficient for the validity of the model.

An empirical test of utility independence requires that statistical criteria be used to decide whether a given set of preference judgments satisfy utility independence. We will describe a simple statistical test of utility independence based on the analysis of variance (ANOVA). To test utility independence, one must elicit certainty matching judgments for gambles of the form $[(Y_1, Q), .5, (Y_2, Q)]$. For any given subject and gamble, there is a hypothetical population of certainty matching judgments produced by that subject for that gamble. We will interpret utility independence statistically as the

---

[3] A proof of Theorem 1 is available on request from John M. Miyamoto.

hypothesis that the population means of certainty matching judgments are unaffected by assumed health state. Thus, if (Y_3, Q) is elicited as the match of $[(Y_1, Q), .5, (Y_2, Q)]$, then utility independence will be interpreted as the hypothesis that the population mean of Y_3 is the same for any choice of Q. So interpreted, utility independence has a straightforward test in an ANOVA. Suppose a set of stimulus gambles is constructed by crossing a set of health states with a set of even-chance gambles for survival duration. If the population means of a subject's certainty matches are the same for every assumed health state, the subject's matching judgments must satisfy the following ANOVA hypotheses:

Hypothesis 1. There is no main effect for the health state.

Hypothesis 2. The (null) effect of health state is the same for different gambles, that is, health state and gamble do not interact.

Conversely, if these two hypotheses are true for every possible even-chance gamble, the health states in question satisfy utility independence. Of course, failure to reject Hypotheses 1 and 2 supports the utility independence assumption only to the extent that the tests have sufficient power to detect violations. Methods for evaluating the power of an ANOVA are well-known and will not be discussed here (Hays, 1973; Scheffé, 1959). The experimental analysis described later includes an analysis of statistical power that illustrates this issue.

Indifference to Health Quality at Short Durations

Before describing our experiment, we must mention a potential violation of the multiplicative utility model that affected both the design and interpretation of the experiment. To describe this potential violation, we must first describe a kind of judgment called a *time trade-off judgment*. Suppose that Q_1 is preferred to Q_2 and that both health states are regarded as better than death. In a time trade-off judgment, an individual is presented with an outcome (Y_2, Q_2), and is asked to state a duration Y_1 such that (Y_1, Q_1) is equal in value to (Y_2, Q_2) (McNeil et al., 1981; Torrance, Thomas, & Sackett, 1972). For example, suppose that (Y_2, Q_2) is a 20-year survival with a specific level of pain, and Q_1 is freedom from pain. Y_1 is the response to the question, "How many years of survival free from pain would be equal in value to 20 years with the specific level of pain?" Responses could range from nearly 0 years to 20 years, depending on the individual's attitude toward the pain.

Utility functions as in Figure 1 are implied if the individual is willing to trade off time in exchange for superior health when survival in the inferior state is long, but not when it is short. For example, a patient might say that 15 years free from pain is equal in value to 20 years with severe pain, but 1 year with severe pain is preferred to any shorter duration free from pain, because survival duration is more precious when it is short. As a matter of terminology, we will say that an individual is *indifferent to health quality at short durations* or *short-term indifferent* if he or she is willing to trade off time for superior health at long durations but not at short durations. Subjects who exhibit this pattern of preference will be called *short-term indifferent subjects.* Pauker (1976) proposed that coronary heart disease patients would be short-term indifferent with respect to survival with and without angina pectoris (chest pain). McNeil et al. (1981) found that many individuals were short-term indifferent in a utility analysis of laryngeal cancer therapy (McNeil et al., 1981). In this study, the superior health state was survival with normal speech, and the inferior health state was survival with impaired speech due to laryngectomy.

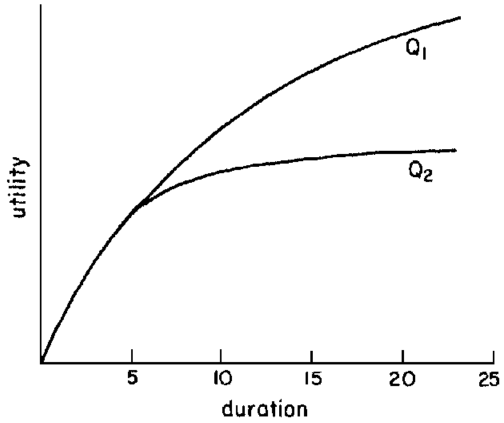

Figure 1. Utility function indicating indifference to health quality at short durations.

In a multiplicative utility model, the utility of survival in different health states must be a constant proportion of each other. In other words, $U(Y, Q_1)/U(Y, Q_2) = F(Y)G(Q_1)/F(Y)G(Q_2) = G(Q_1)/G(Q_2)$, thus proving that the ratio of the Q_1 utility function to the Q_2 utility function is independent of the duration Y. Utility functions like those in Figure 1 imply that the multiplicative model is invalid because the ratio of the Q_1 to Q_2 utility functions is 1.0 at short durations, but it is greater than 1.0 at long durations. Assuming Postulates 1 and 2, it can be proved that any short-term indifferent individual must violate utility independence. Furthermore, the violations must take a specific form: If an individual is short-term indifferent, there must exist gambles that have lower certainty matches when the inferior health state is assumed. (See Proof 2 of the Appendix). This prediction is independent of the specific shape of the utility functions, other than the utility functions are identical at short durations and diverge at longer durations.

We initially pointed out reasons why a multiplicative utility model is plausible, but short-term indifference is a preference behavior that is inconsistent with the multiplicative model. Therefore, we predict that individuals who are not short-term indifferent will satisfy utility independence, and hence, the multiplicative utility model, but short-term indifferent individuals should violate utility independence in the predicted direction, namely, their certainty matches should be lower when the inferior health state is assumed.

Experimental Procedure

Subject Selection

Subjects were inpatients at the Ann Arbor Veterans Administration Medical Center and the University of Michigan Hospital. The sample included patients with cancer, heart disease, diabetes, arthritis and other serious ailments. Patients were contacted by an interviewer who described the experiment as a study of patient preferences for survival risks and health qualities. Patients who were willing to participate in the study were screened according to the following criteria.

First, subjects had to be between 20 and 60 years old. The lower limit of 20 years was simply a crude means to select subjects with a mature outlook on issues of survival. The upper limit of 50 years was chosen because the outcomes in the stimulus gambles ranged from 0 to 24 years. If patients older than 50 years had been sampled, there would be some question whether they regarded 24 years of survival as desirable. It was plausible that patients under 50 years would regard the entire 24 years as desirable.

Each subject was asked to name the health symptoms that impaired his or her quality of life, and to recall the nature of these symptoms during the month preceding the interview. "Current symptoms" was defined to be health symptoms at the severity and frequency experienced by the subject during that month. For example, if the subject had experienced pain and occasional nausea during the month preceding the interview, "current symptoms" referred to pain and nausea at the severity and frequency experienced during that month. The subject and interviewer discussed the concept of current symptoms until the health symptoms comprising its definition had been stated and recorded. Subjects were also asked to consider a second health state, referred to as "freedom from current symptoms," defined to be survival without current symptoms. In this essay, we will often refer to "current symptoms" and "freedom from current symptoms" as "poor health" and "good health." The latter expressions were never used while interviewing subjects.

Two additional screening criteria were formulated using the concepts of survival with and without current symptoms. Subjects were asked whether they would want to live an entire 24 years with current symptoms, or whether a shorter duration was preferable if current symptoms prevailed. If a patient preferred a shorter duration to 24 years with current symptoms, the patient was not included in the sample. Because the sample was restricted to subjects whose preferences for survival were monotonically increasing in the range of 0 to 24 years, the experiment tested only the validity of utility independence within this subpopulation.

Finally, patients were asked to state what duration free from current symptoms would be equal in value to 25 years with current symptoms. The last screening criterion was that the duration elicited by this question had to be less than or equal to 24 years. This criterion was imposed to establish that subjects regarded current symptoms as serious health impairments. If current symptoms were a trivial health impairment, for example, a mild headache, a demonstration that certainty matches are the same assuming survival with and without current symptoms would hardly establish the validity of utility independence. The last criterion established that the current symptoms of patients in the sample were sufficiently serious for them to be willing to give up at least 1 year out of 25 years to be free from current symptoms.

Training in Certainty Matching

Subjects were trained to produce certainty matching judgments as follows. A subject was shown a figure like the one in Figure 2. The diagram on the right stands for an even-chance gamble between 2 and 24 years of survival. The interviewer wrote a number of years, for example, 4, in the open box on the left. The subject was asked to regard this number as a certain survival, and to choose between the certain survival and the gamble. Initially, the subject was told to assume that all survivals were free from current symptoms. After the subject had stated his or her choice, the interviewer erased the certain survival, and wrote a new value in the box. Again, the subject was asked to choose between the certain survival and the gamble.

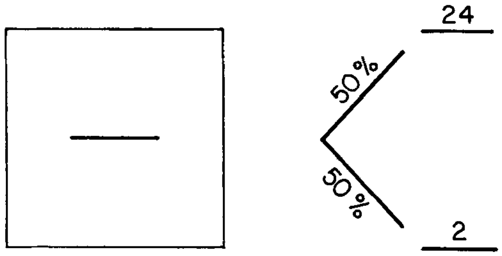

Figure 2. Diagram of a (unspecified) certain outcome and an even-chance gamble as displayed to a subject.

After making a number of choices in this manner, the subject was told that the primary experimental task was not to judge preference between certain outcomes and gambles, but rather to identify a "break-even point," defined to be a duration of certain survival equal in subjective value to the gamble. The subject was told that, by definition, the gamble was preferable to any certain survival less than the break-even point, but less preferred than any certain survival greater than the break-even point. The break-even point was illustrated concretely, using the subject's own judgments with respect to practice gambles. This method of explaining certainty matching, namely, presenting a series of paired comparison choices and then introducing the concept of a break-even point, was found to communicate the task successfully.

The subject was asked to bear in mind the following assumptions when judging certainty matches. First, all survivals used in any single matching judgment would be accompanied by the same health state. Initially, health state was specified to be survival free from current symptoms. At other points in the experiment, it was specified to be survival with current symptoms. Second, any survival duration should be regarded as a survival of the stated number of years, but not more than that many years. Third, the subject was instructed not to regard the certain survival as yielding a situation in which the subject knew that his or her true survival duration would be. The purpose of this instruction was to avoid the following line of reasoning. When considering whether a certain survival or a gamble is preferable, some subjects remark that a short, certain survival has an advantage over a gamble in that one could compress many important experiences into a certain survival, *given that one knew that this was one's remaining lifetime.* This reasoning assumes that certain survivals are qualitatively different from gamble outcomes because they have the added benefit of a known duration. It has the effect of inflating the utility of the certain outcome over that of the gamble outcomes. To avoid this line of reasoning, subjects were instructed to assume that they would not learn what outcome they would receive after having stated a certainty match. This is a natural assumption because it merely recognizes the real situation of the experiment: the experimenter is able to ask the subject to express preferences but is unable to inform the subject as to which outcome will truly occur.

Subjects were taught to bracket the matching value prior to choosing the match, that is, they were instructed to find the match by alternatively considering what values of certain survival were clearly too low or too high. The purpose of the bracketing strategy was to reduce anchoring and adjustment biases that can be present in subjective estimation (Tversky & Kahneman, 1974).

Aspiration Level of Survival

In order to interpret experimental results using prospect theory, it was necessary to determine a reference level for each subject relative to which survival outcomes were regarded as gains or losses. The theoretical role of the reference level will be explained in the Results section, where prospect theory will be described in greater detail. The present study attempted to identify reference levels by defining the concept for subjects, and asking them to judge their own reference levels. When discussing the concept with subjects, the reference level was called the "aspiration level for survival." The concept was explained to subjects as follows:

> I'm going to ask you about something called the aspiration level for survival. Since this concept is fairly complicated, I'll explain it in several steps. The aspiration level for survival is defined to be the length of survival that marks the boundary between those survivals that you regard as a loss and those survivals that you regard as a gain.
>
> For example, my own aspiration level for survival is about the age of sixty. This means that if I found out that I were going to live to the age or fifty or fifty-five (but no more), I would regard this as something of a loss. If I found out that I were going to live to sixty-five or seventy, I would regard this as something of a gain. The aspiration level for survival is not the same as my life expectancy, since my life expectancy is greater than sixty. It's also not the length of time I would want to live, since if I were in good health, I would want to live at least to eighty. The age of sixty is simply a target that marks the boundary between survivals that I would regard to some degree as a loss and survivals that I would regard to some degree as a gain.
>
> I should mention that there's nothing special about the age of sixty. Some individuals place their aspiration level at a very large number, like ninety. For such a person, any survival less than the age of ninety would be regarded to some degree as a loss. I've also encountered individuals who set their aspiration level for survival at their present age. This does not mean that they no longer want to live. It means that they regard every year of survival as a gain. If such an individual learned that he had two years to live, he would regard this as gaining two years of survival, rather than to emphasize some longer survival of which he's being deprived.
>
> Does this concept of an aspiration level of survival make sense to you? Can you tell me what your own aspiration level for survival is?

Subjects generally found these instructions meaningful and would state a value for the reference level without appearing to be confused or uncertain. If a subject did not find the explanation clear, the reference level could be found through a series of concrete questions of the form, "If you found out you were going to live ___years, would you regard this as a loss or a gain?" Subjects found these questions meaningful. Having recognized that survivals could be classified as losses or gains, it became possible to identify an age at the boundary between losses and gains.

Design

Two slightly different experimental designs were used. In the *replicated-judgment* design, each subject produced two certainty matching judgments for each stimulus gamble. In the *unreplicated-judgment* design, certainty matching judgments of gambles were not replicated within subject. Replicating judgments within subject permits testing of hypotheses separately within each subject's data. Data from the unreplicated-judgment design can be used in group analyses, but not in individual subject tests of hypotheses. Different versions of the experiment were developed because it was difficult to recruit subjects for the longer interview necessitated by replicated judgments.

We will first describe the replicated judgment design. Subjects were randomly assigned to one of two conditions, a power condition or an exponential condition. The stimulus gambles used in these conditions were constructed to test the hypotheses that utility functions are, respectively, power or exponential functions of duration. These hypotheses will be examined in a different publication and will not be discussed here. Stimulus gambles for the power and exponential conditions are listed in Table 1. Each stimulus set consists of 6 even-chance gambles for survival duration.

Subjects were asked to judge certainty matches for every gamble of a stimulus set under the assumption of one health state, and then under the assumption of the other health state. Each set of six certainty matching judgments constituted a block, with the same health state assumed to prevail throughout the block. Approximately half of the subjects initially judged certainty matches under the assumption of good health, and subsequently under the assumption of poor health. For the remaining subjects, the trials assuming poor health preceded the trials assuming good health. A third and fourth block of trials replicated the judgments in the first and second blocks. The order of health states was the same in the third and fourth blocks as in the first and second blocks. Stimulus gambles were presented in a different random order in each block.

Subjects were asked to make three time trade-off judgments after each block of certainty matching trials. Subjects judged time trade-offs at 1, 15, and 20 years with current symptoms after the first and third blocks, and at 2, 16 and 24 years with current symptoms after the second and fourth blocks.

In summary, within any subject the replicated judgment design was a 6 × 2 factorial (Gamble × Health State) with two replications per cell. There were two between-subjects variables, power versus exponential conditions, and presentation order for health states. These variables will be called *parametric model* and *presentation order*, respectively.

In the unreplicated judgment design, only the first and second blocks of trials were administered to subjects. Time trade-off judgments were elicited between the first and second blocks at 1, 12, and 20 years with current symptoms. Each subject was classified as short-term indifferent or not short-term indifferent, depending on whether he or she was willing to trade off any time for symptom relief at 1 year with current symptoms. Subjects who were not short-term indifferent proceeded directly to the second block of trials, but short-term indifferent subjects were asked two additional questions. First, they were asked to state a minimum duration at which they would be willing to trade off at least 1 week in order to have survival free from current symptoms. Second, they were asked whether 1 year with current symptoms and 1 year without current symptoms were equally desirable, or whether one outcome was preferable to the other. These questions were introduced in order to settle issues that had been overlooked in the replicated judgment design. The first question attempted to determine the point of divergence between the utility functions for survival in good health and poor health. The second question checked whether the short-term indifferent subjects actually had a preference between health qualities at short durations, even though they were unwilling to trade duration for improved quality.

Finally, the replicated and unreplicated judgment designs differed in the interviewers who conducted the experiment. In the replicated judgment design, one interviewer (Susan Baker) recruited subjects and trained them, and a second interviewer (John Miyamoto) elicited the certainty matching data and time trade-off judgments. In the unreplicated judgment design, a single interviewer (Susan Baker)

Table 1
Stimulus Gambles for Survival Duration

Power condition	Exponential condition
(0, .5, 12)	(0, .5, 12)
(1, .5, 12)	(1, .5, 12)
(4, .5, 12)	(4, .5, 12)
(0, .5, 24)	(12, .5, 24)
(2, .5, 24)	(13, .5, 24)
(8, .5, 24)	(16, .5, 24)

conducted all aspects of the procedure.[4] Criteria for subject selection, training, and elicitation of reference levels for survival duration were identical in the replicated and unreplicated judgment designs.

In summary, the unreplicated judgment design is described by the same factors as the replicated design, except that no replications of certainty matches were elicited within subjects. Replicated and unreplicated judgment subjects differed in the time trade-off judgments elicited between blocks of trials, and in the interviewer who elicited the certainty matching judgments.

Results and Discussion

Proportional Time Trade-Offs

In order to test utility independence in a meaningful way, we must first establish that subjects regard survival with and without current symptoms ("poor health" and "good health") to be clearly different in subjective value. To establish this point, we examine the time trade-off judgments. A time trade-off judgment is a duration X such that X years in good health are judged equal in value to Y years in poor health. The ratio X/Y is called the proportional time trade-off. Each such ratio is a measure of perceived severity of current symptoms because X/Y is small when the individual is willing to give up a large proportion of survival in order to have good health. Table 2 presents mean proportional time trade-offs in various subsamples.

Among replicated judgment subjects ($n = 27$), proportional time trade-offs were averaged over trade off judgments at 15, 16, 20, and 24 years with current symptoms. The mean trade-off in this sample was 71%. Among unreplicated judgment subjects ($n = 37$), proportional time trade-offs were averaged over judgments at 12 and 20 years with current symptoms. The mean trade-off in this sample was 75%. On the average, subjects in the two samples were willing to forego 25% to 30% of long-term survival to be free from current symptoms. These findings support the premise that current symptoms and freedom from current symptoms were sufficiently dissimilar in value to constitute a meaningful test of utility independence.

Transformation of Certainty Matches to Proportional Matches

Prior to performing the ANOVA, it is useful to transform each certainty match to a new quantity called a *proportional match*. To define this transformation, let CM denote the judged certainty match of a gamble between a higher duration (HIGH) and a lower duration (LOW). Define a new quantity PM by the following rule:

$$PM = \frac{CM - \text{LOW}}{\text{HIGH} - \text{LOW}}.$$

The number PM will be called the proportional match corresponding to the certainty match CM.

Table 2
Age and Proportional Time Trade-Offs

Subjects	n	Age M	Age SD	% Long-term survival[a,b] M	% Long-term survival[a,b] SD
		Replicated judgment			
Non-short-term indifferent	20	33.7	7.2	71	17
Short-term indifferent	7	34.0	4.4	72	09
All	27	33.8	6.5	71	15
		Unreplicated judgment			
Non-short-term indifferent	27	33.5	9.3	74	15
Short-term indifferent	10	31.9	8.1	79	26
All	37	33.1	9.0	75	18

[a] Percentages represent proportions X/Y, where X years in good health is judged equal in value to Y years in poor health. [b] Each replicated judgment subject's mean was averaged over time trade-offs at 15, 16, 20, and 24 years in poor health. Each unreplicated judgment subject's mean was averaged over time trade-offs at 12 and 20 years in poor health.

Utility independence predicts that there is no main effect of health quality (Hypothesis 1) and no interaction between health quality and gamble (Hypothesis 2). Because each PM is linearly related to a corresponding CM, certainty matches satisfy Hypotheses 1 and 2 if and only if proportional matches satisfy Hypotheses 1 and 2. Hence, we can test Hypotheses 1 and 2 by using proportional matches as data instead of certainty matches.

There are two advantages to using proportional matches as data. First, estimates of variance within cell generally increase as a function of the range of the stimulus gamble. The transformation to proportional matches tends to equalize the variance of responses to different gambles. Second, proportional matches are more informative than certainty matches, especially when results are presented as averages across gambles. For example, if a subject's mean proportional match is .4, one knows that his or her average response was slightly below the expected value of a gamble, but if the mean certainty match is 9.3 years, one cannot interpret this result without knowing the stimulus gambles to which it was a response.

The Effect of Health State: Individual Subject Analysis

A separate ANOVA based on proportional matches was calculated for each subject in the replicated judgment sample. The F statistic for health quality was significant at the .05 level for 6 of 27 subjects. If the null hypothesis truly described every subject, the probability would be 95.6% that three or fewer F statistics would be significant out of 27 F tests at the .05 level.[5] It is clear that some subjects violate utility inde-

---

[4] We would like to thank Sue Baker for recruiting and training subjects, and administering the experimental procedure to the unreplicated-judgment subjects.

[5] The probability of 3 or fewer significant Fs assuming that the null hypothesis was true in all 27 cases was calculated using the cumulative binomial for an event with a .05 probability of occurrence.

pendence, but further analyses are needed to determine whether they are few in number as this initial finding suggests.

Figure 3 displays a scatter plot of mean proportional matches for the 27 subjects. Each point represents one subject's mean proportional match in the good health condition versus poor health condition. If utility independence were true of every subject, the means would lie close to the diagonal with deviation from the diagonal due only to random variation. As can be seen, the means for most subjects lie close to the diagonal. The absolute value of the mean difference between good health and poor health matches was less than .05 for 16 subjects, between .05 and .1 for 7 subjects, and greater than .1 for 4 subjects. Thus, the average difference between good health and poor health matches was less than 1/20th of the range of gambles for 16 subjects, and between 1/10th and 1/20th of the range for another 7 subjects. The mean differences were generally quite small.

In Figure 3, an open circle or triangle indicates a pair of means whose difference was significant at the .05 level. A close circle or triangle indicates a pair of means whose difference was not significant. The distinction between circles and triangles pertains to prospect theory, and will be explained later. Of 6 subjects with significant *F*s, 5 had larger mean proportional matches assuming poor health ($p = .22$, by a sign test). Of 21 subjects with nonsignificant *F*s, 13 had larger mean proportional matches assuming poor health ($p = .38$, by a sign test). Although the pattern in Figure 3 suggests a slight tendency for mean proportional matches to be larger assuming poor health, with the tendency concentrated among subjects with significant *F*s, the apparent tendency is not statistically significant.

Tests of the interaction between health quality and gamble yielded 2 significant *F*s at the .05 level. Neither subject with a significant interaction had a significant main effect. If the hypothesis of no interaction were true of all 27 subjects, there would be a 39.4% chance of observing two or more significant *F*s at the .05 level.[6] Thus, the results are consistent with the hypothesis that health quality does not interact with stimulus gamble.

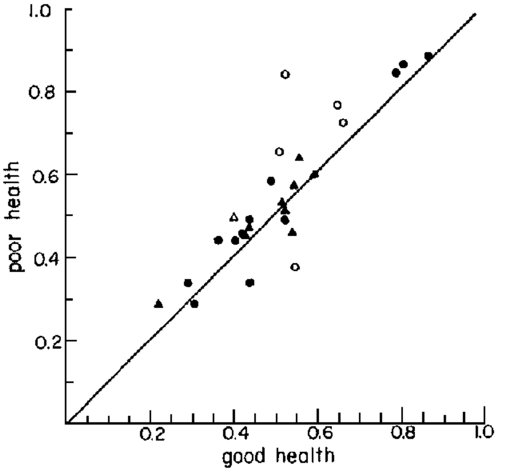

Figure 3. Scatter plot of mean proportional matches in the good health and poor health conditions.

Power calculations are required to determine whether the low frequency with which utility independence was violated supports the conclusion that most subjects satisfy this property. To describe the power analysis, let μ_{DIF} denote the population difference between a subject's mean proportional matches in the good health and poor health conditions. Let σ^2 denote the population variance of a proportional match (assuming that the variance for different gambles is the same within any subject), and let MS_e denote the mean squared error calculated from the subject's data. The power of the *F* test can be approximated if one knows or assumes specific values for μ_{DIF} and σ^2 (Hays, 1973, Section 10.18). We made the heuristic assumption that each subject's variance equaled his or her mean squared error. (MS_e is an unbiased estimator of σ^2.) Given this assumption, we calculated each subject's probability of rejecting the hypothesis of no effect of health quality at the .05 level when the true value of μ_{DIF} was equal to .05, .1, and .2.

Figure 4 displays cumulative sample distributions of the power of individual subjects' tests against various alternatives. To understand this graph, consider the curve on the left (labeled $\mu_{DIF} = .05$). For each subject, the probability of rejecting the null hypothesis at the .05 level was calculated given that μ_{DIF} equaled .05 and σ^2 equaled MS_e. The curve labeled .05 is the cumulative sample distribution of these probabilities, that is, the ordinate represents the proportion of the sample whose power was less than or equal to the value on the absissa. For example, given $\mu_{DIF} = .05$ and $\sigma^2 = MS_e$, the probability of rejecting the null hypothesis was less than .2 for 44% of the sample (12 subjects), and less than .6 for 81% of the sample (22 subjects). The curves in the middle and on the right are the cumulative sample distributions of power against μ_{DIF} equal to .1 and .2.

Clearly the likelihood of detecting a true difference of .2 was quite high for all subjects. The power against this difference exceeded .8 for all subjects, and exceeded .9 for 26 of 27 subjects.

The issue for a true difference of .1 is more complex, but it can be shown that the experiment had considerable power against this alternative as well. Note first that the minimum power against $\mu_{DIF} = .1$ was .28, and the median power was .66. Thus, all subjects had moderate to large probabilities of detecting a .1 difference. To determine whether these probabilities were compatible with the observation of 6 significant *F*s in the sample as a whole, a computer was programmed to calculate the probabilities of 0, 1, 2, . . . , 27 significant *F*s, given the estimates of power against $\mu_{DIF} = .1$.[7] The probability of 6 or fewer significant *F*s was less than .0001, and the probability of 15 or more significant *F*s was greater than .95. If a true difference of .1 had been prevalent in the sample, one would expect many more significant *F*s than were observed.

---

[6] The calculation was based on the cumulative binomial, as in Footnote 5.

[7] The calculation uses a standard combinatorial formula for the probability of *K* events out of *N* independent events, where the events have probabilities $p_1, p_2, \ldots, p_N$ of occurring. (Cf. Feller, 1950).

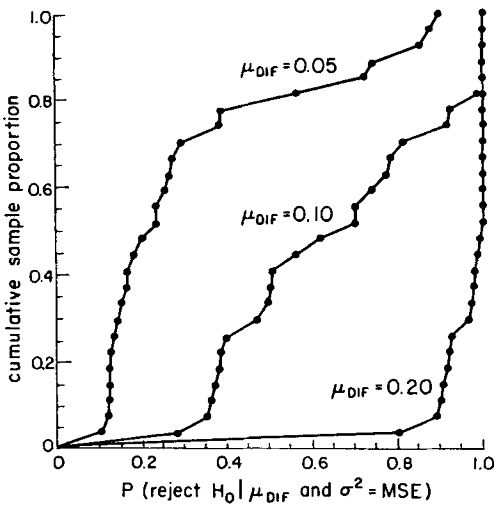

Figure 4. Cumulative sample distributions of power against a true mean difference (μ_{DIF}) of .05, .1 and .2. (Each subject's variance [σ^2] is assumed equal to subject's mean squared error [MS_e]. The null hypothesis is tested at the .05 level of significance.)

The experiment had relatively little power against μ_{DIF} = .05. Given the individual estimates of power against μ_{DIF} = .05, the probability of 6 or fewer significant *F*s was found to be .12. Therefore, the observation of 6 significant *F*s would be somewhat unlikely, but far from impossible, even if every subject had a small effect (μ_{DIF} = .05) for health quality.

The power analysis indicates that a true difference of .2 would almost certainly be detected, and a true difference of .1 would often be detected. True differences of .05 or less, however, would be indistinguishable from zero differences in the present experiment. Although, in principle, no results could ever prove that the population means of good health and poor health matches are precisely equal, the power analysis shows that only small effects of health quality are compatible with the experimental results.

The power of the test for Gamble × Health Quality interactions must also be evaluated. Unfortunately, the preceding analysis cannot be repeated with respect to the interaction effects because it was based on an approximating formula for power that only applies when the *F* has 1 degree of freedom in the numerator.[8] Because the *F* for interaction has 5 degrees of freedom in the numerator, we used the Pearson and Hartley charts (reprinted in Scheffé, 1959) to estimate power against interaction effects. Even with these charts, the power against interaction effects could not be estimated for every subject because many of the values that were needed were lower than the values presented in the charts. Therefore, we will discuss power calculations for the median subject only.

The median MS_e was .0079. If one assumes that the true interaction effects were ±.10 for individual gambles, then the probability of detecting this interaction would be .68. If the true interaction effects were ±.05 for individual gambles, then the probability of detecting the interaction is lower than the lowest value on the Pearson and Hartley chart for 5 and 12 degrees of freedom and the .05 significance level. Visual extrapolation from the chart makes it reasonably clear that the power must be between .15 and .25 (best guess = .20). The power analysis shows that a true interaction of ±.10 would have been detected by many subjects, but for moderate to small interactions (less than ±.05), the chances of detecting the interaction were rather low.

The Effect of Health Quality: Group Analysis

Separate analyses of group data were carried out for the replicated judgment sample, the unreplicated judgment sample, and the two samples combined. These analyses revealed essentially the same picture, so only the analysis of the combined samples will be reported here.

With respect to Blocks 1 and 2, the replicated- and unreplicated-judgment designs differed only in the interviewers who elicited the certainty matching data, and in the time trade-off questions intervening between blocks. Therefore, data from Blocks 1 and 2 of replicated and unreplicated judgment subjects were combined, and data from Blocks 3 and 4 of replicated judgment subjects were discarded. The ANOVA for the combined samples had health quality and gamble as within-subjects variables; interviewer, parametric model, and presentation order as between-subjects variables; and subjects as a random-effects factor nested within interviewer, parametric model, and presentation order. Certainty matches were transformed to proportional matches prior to statistical analysis.

Health state did not have a significant main effect on proportional matches, $F(1, 56) = 2.42$, $MS_e = .039$, $p > .10$. The mean difference between the good health and poor health conditions was −.021 ± .029 with 95% confidence. When averaged across subjects and gambles, matches in the good health and poor health conditions differed at most by a small proportion of the range of a gamble. The interaction between health quality and stimulus gamble was also not significant, $F(5, 280) = 1.56$, $MS_e = .011$, $p > .10$.

The results of a sign test show that there is actually a slight tendency for proportional matches to be greater in the poor health condition. The data from Blocks 1–4 of the replicated-judgment subjects show that 18 of 27 had larger mean proportional matches in the poor health condition. The data from Blocks 1 and 2 of the unreplicated-judgment subjects show that 23 of 37 subjects had larger mean proportional matches in the poor health condition. Combining these results yields 41 of 64 subjects with larger mean proportional matches in the poor health condition ($z = 2.13$, $p < .05$). Recall from the individual subject tests that the tendency to produce larger proportional matches in the poor health condition was largely confined to subjects with significant *F*s for health quality. Hence, the results are consistent with the interpretation that a small proportion of subjects systematically produce larger proportional matches in the poor health condition, whereas the majority of subjects have no underlying effect of health quality.

---

[8] The formula that we use is given in Hays (1973, Section 10.18) and Scheffé (1959, p. 415).

Indifference to Health Quality at Short Durations

Earlier, we derived the prediction that short-term indifferent subjects should have lower certainty matches assuming poor health. The test of this prediction is slightly different for replicated- and unreplicated-judgment subjects. The replicated judgment subjects will be discussed first.

Replicated-judgment subjects were not asked directly whether they had preferences between health qualities at short durations. Therefore, an operational definition was adopted according to which subjects were classified as short-term indifferent if they did not trade off any time for symptom relief on any trial at 1 or 2 years with current symptoms. (There were four such trials). Seven subjects were classified as short-term indifferent under this criterion. Note that these subjects did regard their current symptoms as serious impairments. On the average, they were willing to forego 28% of long-term survival to achieve symptom relief (see Table 2). Thus, these subjects were indifferent to health quality at 1 and 2 years of survival, but were willing to trade off substantial proportions of long-term survival for improved health.

Only 1 short-term indifferent subject had a significant main effect of health quality ($p < .05$, two-tailed) and 1 other had a significant interaction between health quality and gamble ($p < .01$, two-tailed). Both subjects had higher mean proportional matches in the poor health condition. The same pattern also held for 3 of the 5 remaining subjects. Thus, 5 of 7 subjects had higher mean proportional matches in the poor health condition, contrary to the implications of short-term indifference and Postulates 1 and 2.

The results for unreplicated judgment subjects were qualitatively similar. Unreplicated judgment subjects were classified as short-term indifferent if they were willing to trade off any time for symptom improvement when survival with current symptoms was 1 year. Ten unreplicated-judgment subjects satisfied this criterion. Of the 10, 7 had higher mean proportional matches in the poor health condition, contrary to the predicted pattern for short-term indifferent subjects.

Combining the results for replicated and unreplicated judgment subjects, we found that 12 of 17 subjects had higher mean proportional matches in the poor health condition ($p = .14$, by a sign test). An ANOVA was computed for the combined sample of short-term indifferent subjects. As before, only Blocks 1 and 2 of the replicated judgment subjects were used in the analysis. The main effect of health quality was not significant, $F(1, 9) = 1.44$, $MS_e = .051$, $p > .25$. Mean proportional matches were $.037 \pm .070$ greater in the poor health condition with 95% confidence. The interaction of health quality with gamble was also not significant, $F(5, 45) = .53$, $MS_e = .014$, $p > .75$. Because the confidence interval for the mean difference between poor health and good health matches includes negative values, we cannot exclude the possibility that the true difference is in the predicted direction (i.e., good health exceeds poor health). If the predicted effect exists, however, it must be quite small.

One might object that indifference to health quality at short durations and Postulates 1 and 2 merely predict that the certainty matches of *some* gambles must be lower when poor health is assumed; this does not preclude the existence of other gambles with higher certainty matches assuming poor health. If there were enough such gambles, mean proportional matches of short-term indifferent subjects could be higher in the poor health condition without contradicting Postulates 1 and 2. Although this ad hoc assumption would explain why mean proportional matches of short-term indifferent subjects were generally greater in the poor health condition, it would also predict that the effect of health quality should interact with stimulus gamble. A significant interaction was found in only one short-term indifferent subject in the replicated-judgment sample, and the interaction in the group data did not approach significance.

In summary, Postulates 1 and 2 imply that for any short-term indifferent subject, there exist gambles that have lower certainty matches when poor health is assumed (Proof 2 of the Appendix). Although the experimental results do not conclusively reject the prediction, the data are qualitatively opposite to it. Thus, the results suggest that short-term indifferent subjects do not satisfy both Postulates 1 and 2. In particular, the results suggest that preferences for riskless options are inconsistent with preferences for options that involve risk. To see this, note that Postulate 1 implies that the utility function represents the preference equivalencies expressed in time trade-off judgments, a riskless preference relation. Postulate 2 implies that the utility function also represents the subjective averaging that underlies choice of the certainty match, a preference relation between a riskless and a risky option. The assumption that the same utility function mediates preferences for riskless options and risky options implies that certainty matching judgments must be consistent with time trade-off judgments in the predicted way. This prediction was not confirmed.

A conclusive demonstration that short-term indifferent subjects have equal or higher certainty matches assuming poor health would directly contradict the widely held assumption that the subjective values of outcomes are the same, regardless of whether one is judging preference between riskless options, or preference between gambles for these outcomes. Stated in terms of the utility model, the results appear to contradict the assumption that the same utility function mediates preferences for riskless and risky options. Although the present evidence is not conclusive, it is important simply because it makes the existence of such a contradiction plausible. The possibility of this inconsistency has not previously been recognized in the theory of preference behavior.

Lexicographic Preferences at Short Durations

Some short-term indifferent subjects exhibited a form of preference behavior called a lexicographic preference order (Coombs, 1964; Tversky, 1969). Preferences for survival duration and health quality are said to be lexicographically ordered if preference is determined exclusively by duration whenever outcomes differ in duration, but preference is determined exclusively by quality when the durations of outcomes are equal. For example, if a patient's preferences were lexicographic, then the patient would prefer 2 years with pain to any shorter duration without pain, but he would prefer 2 years without pain to 2 years with pain.

In the unreplicated judgment sample, 3 of 10 short-term indifferent subjects had lexicographic preferences at short durations. They were unwilling to trade 1 year with current symptoms for any shorter duration without current symptoms, but they preferred 1 year without current symptoms to 1 year with current symptoms. The remaining 7 subjects said that either health quality was equally desirable if survival lasted only 1 year. Although this judgment may seem peculiar to those who do not share these values, such subjects justify their preference with remarks to the effect that when survival duration is short, health quality is unimportant. We could not determine whether short-term indifferent subjects in the replicated judgment sample had lexicographic preferences because we neglected to ask them whether they had a preference between a short survival with and without current symptoms. No systematic differences were found between the short-term indifferent subjects who were or were not lexicographic at short durations.

Analysis in Terms of Prospect Theory

To apply Kahneman & Tversky's (1979) prospect theory to the present experiment, we need to explain the concept of a reference level of survival duration and a distinction between regular and irregular gambles. These terms are needed because prospect theory requires that different tests of the multiplicative utility model be carried out for the regular and irregular gambles, and these classes of gambles are defined in terms of the reference level of survival duration. Although the present discussion of prospect theory is self-contained, the motivation for prospect theoretic formulations will not be adequately covered (see Kahneman & Tversky, 1979, 1984; Miyamoto, 1987). Readers who are unfamiliar with prospect theory can skip this section without loss of continuity.

Prospect theory postulates that an individual has a subjective reference level that is regarded as neutral or status quo. Outcomes are perceived as gains if they are preferable to the reference level, and as losses if they are less preferred than the reference level (Kahneman & Tversky, 1979, 1984; Tversky & Kahneman, 1974). For example, when betting with money, the zero outcome is a natural reference level, with money received in the bet perceived as a gain, and money paid out perceived as a loss. In applying prospect theory to gambles for survival duration, we assumed that subjects could judge whether a given survival duration was a loss or a gain. Each subject was asked to identify his or her own reference level, defined as the boundary between those survivals that were regarded as losses and those survivals that were regarded as gains.

Prospect theory distinguishes between two kinds of gambles, called regular and irregular gambles. We will not give the full definition of these gambles, but only define them with respect to even-chance gambles. An even-chance gamble is regular if one of its outcomes is equal or greater than the reference level, and the other outcome is equal or less than the reference level. An even-chance gamble is irregular if both outcomes are greater than the reference level or both are less than the reference level. For example, if the reference level of an individual is 10 years of survival, then (1 year, .5, 20 years) and (10 years, .5, 25 years) are regular gambles, and (1 year, .5, 9 years) and (11 years, .5, 20 years) are irregular gambles.

The distinction between regular and irregular gambles is relevant to the test of the multiplicative utility model because prospect theory implies that Postulate 2 holds separately in the regular gambles and in the irregular gambles, but not necessarily in the combined set of all even-chance gambles. The reason for this is that prospect theory allows for the possibility that s and t, the weights in Postulate 2, differ for regular and irregular gambles. Because the test of the multiplicative utility model assumes that Postulate 2 is valid with respect to a single choice of s and t, the interpretation of results from the standpoint of prospect theory must be carried out separately for regular and irregular gambles.

Although the distinction between regular and irregular gambles makes the prospect theoretic interpretation of results rather complicated, there is a subset of subjects for whom the analysis is straightforward. The survival durations in the stimulus gambles of the present experiment ranged from 0 to 24 years. Hence, any subject whose reference level exceeded 24 years is predicted to categorize every stimulus gamble as irregular. Such subjects will be called purely irregular subjects because all stimulus gambles were irregular relative to their reference levels. Prospect theory implies that purely irregular subjects satisfy Postulate 2 with respect to a single choice of s and t for every stimulus gamble of this experiment. Because Postulates 1, 3, and 4 remain plausible in the context of prospect theory, we can test utility independence in the data from purely irregular subjects, and the tests can be interpreted from the standpoint of prospect theory, just as we previously interpreted the tests from the standpoint of the dual bilinear model, expected utility theory and subjective expected utility theory. Subjects who were not purely irregular will be omitted from the analysis because the analysis of their data is much more complex. For these subjects, some stimulus gambles were irregular and others were not, and separate analyses would be required for each subject's idiosyncratic division of stimulus gambles into regular and irregular gambles.

There were 17 purely irregular subjects among the 27 replicated-judgment subjects, and 33 purely irregular subjects among the 37 unreplicated-judgment subjects. Five of the 17 replicated-judgment subjects had significant main effects ($p < .05$) for health quality, and 1 had a significant interaction ($p < .01$) between health quality and gamble. In Figure 3, the mean proportional matches of purely irregular subjects are indicated by circles that are either open or closed depending on whether the F statistic for health quality was significant. The pattern of results for purely irregular subjects does not differ systematically from the pattern for all subjects. An ANOVA was calculated for the 50 purely irregular subjects using proportional matches from Blocks 1 and 2. The main effect of health quality was not significant, $F(1, 42) = 1.54$, $p > .20$. The mean difference between the good health and poor health conditions was $-.025 \pm .040$ with 95% confidence. The interaction between health quality and gamble was also not significant, $F(5, 210) = .82$, $p > .50$. The results for the purely irregular subjects are consistent with the view that most of these subjects satisfy utility independence.

The analysis in terms of prospect theory serves two purposes. First, it shows that if one has a method for determining a subject's reference level, it is possible to test the multiplicative utility model within the prospect theory framework. Miyamoto (in press) has shown how to formulate a number of different multiattribute utility models in the prospect theory framework using an approach that is very similar to the one developed here. Second, we were curious whether the results for purely irregular subjects would differ substantially from the results for the sample as a whole. What we found is that the results for purely irregular subjects were very similar to the results for the entire sample. Thus, the analysis using prospect theory is consistent with our initial analysis, where Postulates 1–4 were assumed of all subjects and the concept of a reference level was ignored.

General Discussion

The Multiplicative Utility Model

In our formalization of the multiplicative utility model, Postulates 1 and 2 were assumed on the basis of prior theory, and Postulates 3 and 4 were taken to be intuitively obvious. Hence Postulate 5, utility independence, emerged as the main empirical hypothesis to be tested. The multiplicative utility model predicts that survival duration is utility independent of health quality. Conversely, if Postulates 1–4 are assumed to be true, then utility independence is sufficient to establish the validity of the multiplicative utility model. The experimental results suggest that most subjects satisfy utility independence. Were it not for the results for short-term indifferent subjects, we would conclude that the utility representation of most non-short-term indifferent subjects is multiplicative. To understand why the results for short-term indifferent subjects affect our conclusions, we must consider the implications of the fact that they did not generally violate utility independence.

Although the failure of short-term indifferent subjects to violate utility independence may merely indicate that their violations were too small to be detected, there are two reasons for doubting this interpretation. First, the experiment did have substantial power to detect effects of health quality at the individual subjects' level. Second, mean proportional matches were larger in the poor health condition for 12 of 17 short-term indifferent subjects, contrary to the prediction that short-term indifferent subjects would produce smaller proportional matches in the poor health condition. It is possible for mean proportional matches of short-term indifferent subjects to be larger in the poor health condition if the effect of health quality interacts with stimulus gamble. There was little evidence for such an interaction. Although the power of the individual subjects' tests for interaction was only sufficient to detect rather large interaction effects, the interaction in the group analysis of short-term indifferent subjects was also not significant. Furthermore, there is no theoretical reason for predicting such an interaction.

The prediction that short-term indifferent subjects should violate utility independence is derived from the assumption that the same utility scale mediates the ordinal preference relations described by Postulate 1 and the subjective averaging described in Postulate 2. The results for short-term indifferent subjects are in conflict with this assumption. Virtually every theory of preference under risk postulates that the worth of gambles is determined by a process of subjective averaging, and subsumes ordinal relations of riskless preference as a special case of preference between gambles (Edwards, 1962; Kahneman & Tversky, 1979; Karmarkar, 1978; Luce & Narens, 1985; Luce & Suppes, 1965; Savage, 1954; Schoemaker, 1982). The present results suggest that, in general, short-term indifferent subjects do not have utility scales that represent both the ordinal relations of riskless preference, and the subjective averaging that underlies judgments of certainty matches. Although the results did not conclusively demonstrate an inconsistency between short-term indifference and certainty matching, the evidence was sufficiently strong to show that the possibility of such an inconsistency deserves serious consideration.

Tversky (1967) reported evidence for inconsistency between riskless and risky measures of utility, but the pattern of results that he found were qualitatively different from the present findings. Tversky's (1967) results suggested that either the utilities measured by preferences between risky options differ systematically from utilities measured by preferences between riskless options, or else the subjective probabilities of complementary events do not sum to one. To compare these results with our own, we should first note that Postulates 1–5 imply the multiplicative utility representation even if $s + t \neq 1$, where s and t are the weights in Postulate 2 (Miyamoto, 1985, in press). Therefore, even if the subjective probabilities of complementary events do not sum to one, Postulates 1–4 and short-term indifference imply that utility independence is violated. If we agree that short-term indifferent subjects satisfy utility independence, or else violate it in the unpredicted direction, we must reject at least one of Postulates 1–4. Given that Postulates 3 and 4 are very plausible, Postulates 1 and 2 are the suspect assumptions. Unlike Tversky's (1967) results, the assumption that the subjective probabilities of complementary events do not sum to one does not eliminate the inconsistency between the risky and riskless preferences of short-term indifferent subjects.

If one assumes instead that the subjective probabilities of complementary events sum to one, then Tversky's results show that quantities of candy and cigarettes have greater utility relative to the utility of money when measured using preferences between gambles than when measured using preferences between riskless options. As Tversky (1967) argued, these results can be explained if one assumes the existence of a positive utility of gambling. However, Tversky's results support the interpretation that riskless and risky utilities are at least monotonically related to each other, whereas the results for short-term indifferent subjects indicate an ordinal inconsistency between riskless and risky utilities. Hence, the implications of short-term indifference for the representation of utility are different from the implications of Tversky's (1967) study.

From a psychological standpoint, it is not hard to see how the integration of stimulus dimensions could be quite different in riskless preference and certainty matching. Paired comparison choice between multiattribute outcomes encourages interdimensional comparisons (Tversky, 1969). When judging preference between (Y_1, Q_1) and (Y_2, Q_2), one must decide whether the difference in duration between Y_1 and Y_2 compensates the difference in quality between Q_1 and Q_2. Paired-comparison choice makes salient the relative worth of durations across health qualities. Thus, it is easy for a subject to adopt a strategy that equates the worth of Y years in health states Q_1 and Q_2 when Y is short, but not when Y is long. Stated more intuitively, the subject can decide that when survival duration is short, symptom relief does not compensate any reduction in duration.

To see why short-term indifference should also affect certainty matching judgments, note that if one is short-term indifferent, utility differences between durations that are less than the point at which the utility functions diverge are the same in different health states, but utility differences between durations that are greater than this point are smaller when poor health is assumed (see Figure 1). This complex implication of short-term indifference is essentially why the certainty matches of some gambles must be lower when poor health is assumed (see Proof 2 of the Appendix).

As opposed to this, we note that the following plausible strategy yields matching judgments that satisfy utility independence. When judging a certainty match, health quality is the same in the matching outcome and the gamble outcomes. Because the stimulus and response vary only on the dimension of survival duration, one can adopt a strategy for certainty matching that ignores health quality and focuses only on survival duration. A strategy of ignoring the common health quality is similar to editing operations in prospect theory (Kahneman & Tversky, 1979). Certainty matches produced by such a strategy will satisfy utility independence, because judgments of matches will not be affected by assumed health state.

The results for short-term indifferent subjects do not directly contradict the hypothesis that subjects who are not short-term indifferent satisfy a multiplicative utility model, but they cast doubt on Postulates 1 and 2, which are assumed in the derivation of the model. One way to remove these doubts would be to test whether subjects who satisfy Postulates 3 and 4 and utility independence also satisfy ordinal preference axioms from which Postulates 1–2 can be derived. Miyamoto (1985, in press) formulated a measurement axiomatization of Postulates 1 and 2, but did not test the axioms empirically. A precise statement of this axiomatization cannot be given here.

In designing the present study, we assumed that Postulates 1 and 2 were secure from prior theory. In the light of the results for short-term indifferent subjects, however, it now appears that the ordinal assumptions underlying Postulates 1 and 2 must be tested along with utility independence before one can conclude that subjects satisfy a multiplicative utility representation. What is clear, however, is that either subjects violate a fundamental assumption of preference theory, namely, mutual consistency of the riskless preference ordering and relations of certainty matching, or else subjects who are not short-term indifferent generally satisfy a multiplicative utility model.

Indifference to Health Quality at Short Durations

Indifference to health quality at short durations has something of the flavor of the lexicographic semiorder that gave rise to intransitivities of preference in Tversky's (1969) classic study. It is therefore interesting to compare the two phenomena. Suppose an individual must choose between alternatives A and B, where these alternatives are described by their values on two dimensions. The preference structure is a lexicographic semiorder if one bases the choice exclusively on Dimension 1 when the alternatives differ by more than ϵ on Dimension I ($\epsilon > 0$), but one bases the choice exclusively on Dimension II when the difference on Dimension I is less than ϵ (Tversky, 1969). For example, if $\epsilon = 2$ years, and one must choose between (Y_1, Q_1) and (Y_2, Q_2), then one ignores quality and chooses the outcome with the longer duration if $| Y_1 = Y_2 | > 2$, but one ignores duration and chooses the outcome with the superior quality if $| Y_1 - Y_2 | \leq 2$.

We should distinguish the lexicographic semiorder from the lexicographic order that was observed in some short-term indifferent subjects. In a lexicographic order, one chooses the outcome with the longer duration if $Y_1 \neq Y_2$, and one chooses the outcome with the superior quality if $Y_1 = Y_2$. In effect, a lexicographic order is simply the special case of a lexicographic semiorder where $\epsilon = 0$. Tversky (1969) showed that if preferences are described by a lexicographic semiorder (but not a lexicographic order), then the preference order will be intransitive in the sense that $A >_p B$, $B >_p C$ and $C >_p A$ for some outcomes A, B and C. If the preference order is lexicographic but not a semiorder (i.e., $\epsilon = 0$), then intransitivities of preference cannot occur.

The decision rule that underlies short-term indifference has a different structure from the lexicographic semiorder. A short-term indifferent individual bases choice on both Dimensions I and II (duration and quality) when values on Dimension I are large in absolute magnitude. When values on Dimension I are small, a short-term indifferent subject is either lexicographic, or else truly indifferent to health quality. Whereas in a lexicographic semiorder, attention switches from Dimension I to Dimension II when *differences* on Dimension I are small, the short-term indifferent subject switches from utilization of both dimensions to utilization of Dimension I exclusively or to a lexicographic ordering of the dimensions when *values* on Dimension I are small *in absolute magnitude*. Thus, indifference to health quality at short durations is only superficially analogous to a lexicographic semiorder.

Indifference to health quality at short durations is more akin to Weber's law in magnitude discrimination, or diminishing marginal value in utility theory. A given difference in stimulus intensity appears greater when the intensities are small. A given difference in money appears greater to a poor man than to a rich man. The analogy is not perfect, because indifference to health quality at short durations is dependent on the use of two dimensional stimuli rather than one dimensional stimuli. One way to describe indifference to health quality at short durations, however, is that differences in survival duration become infinitely large relative to differences in health quality when the absolute magnitude of duration is short. It is the relative magnitude of subjective differences across dimensions, rather than the relative magnitude of subjective differences within a dimension that is affected by the absolute magnitude of survival duration.

Methodological Considerations

One difficulty in the study of utility models is that several different theories of preference under risk have been proposed to account for empirical violations of expected utility theory (Kahneman & Tversky, 1979; Karmarkar, 1978; Luce & Narens, 1985). Because formalization of a specific utility model like the multiplicative model requires that one assume the framework of some theory of preference, one must be concerned with the possibility that the analysis of the model is only interpretable within the background theory that is adopted as framework.

Our formalization and experiment represent one approach to overcoming this difficulty, for they can be interpreted within any theory that implies the validity of Postulates 1-3. Thus, the analysis can be interpreted within nonstandard utility theories (Karmarkar, 1978; Luce & Narens, 1985; Shoemaker, 1982), as well as within expected utility or subjective expected utility theory (Savage, 1954; Von Neumann & Morgenstern, 1953). Furthermore, the results for purely irregular subjects can be interpreted within prospect theory, if one accepts the validity of judgments of survival reference levels. The limited scope of Postulates 1-3 has the advantage that one can study models of utility under these assumptions without commitment to any one of several theories of preference under risk. This point might seem to be undermined by the possibility that some short-term indifferent subjects violate Postulates 1 and 2. In a backhanded way, however, the finding further demonstrates the value of the present formalization, for by limiting the scope of Postulates 1-3, we broaden the class of theories that can be rejected if these postulates are found to be empirically invalid.

Although utility independence is not a fundamental assumption of preference theory, in the sense that violations of utility independence do not undermine the foundations of preference theory, it is nevertheless a basic theoretical assumption in the study of multiattribute utility models. In closing, we will briefly explain why this is the case.

A multiplicative utility model in two attributes was defined to be a utility scale U that satisfies Equation 1. Similarly, an additive utility model in two attributes postulates the existence of separate scales F and G such that

$$U(Y, Q) = F(Y) + G(Q) \tag{12}$$

for every value of Y and Q. (We continue to draw upon the example of survival duration and health quality, although our remarks apply to any two-attribute utility problem.) Assumptions of utility independence are related to additive and multiplicative utility models by the following fundamental result:

Theorem 2: If two attributes are such that the first is utility independent from the second, and the second is utility independent from the first, then the utility scale is either additive or multiplicative, that is, either Equation 1 or 12 holds. Conversely, the additive and multiplicative utility models both predict that each attribute is utility independent from the other.

This result was first proved under the assumption of expected utility theory (Keeney, 1968, in 1971; Keeney & Raiffa, 1976; Raiffa, 1969). Miyamoto (1985, in press) showed that it also holds under Postulates 1-3, which are weaker than the assumptions of expected utility theory.

Note that Theorem 2 does not indicate which of the two models, the additive or multiplicative utility model, is valid. It merely tells us that if two attributes are utility independent from each other, then the only utility models consistent with this relation are the additive and multiplicative models. Some additional diagnostic property must be considered in order to determine which model is valid. Postulate 4 states a diagnostic property that excludes the additive utility model. Other diagnostics that distinguish additive from multiplicative utility models are described in Keeney and Raiffa (1976), Krantz et al. (1971), and Miyamoto (in press).

This is not the place for a full discussion of utility independence assumptions. Theorem 2 is presented here only to make the point that assumptions of utility independence play a central role in the investigation of additive and multiplicative models throughout multiattribute utility theory (Keeney & Raiffa, 1976; Miyamoto, 1983, in press). Consequently, the theoretical and empirical analyses presented here demonstrate the feasibility of experimental investigations of these models in a framework that encompasses nonstandard utility theories like prospect theory and the dual bilinar model, as well as the standard expected utility and subjective expected utility theories.

References

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

Coombs, C. H. (1983). *Psychology and mathematics: An essay on theory.* Ann Arbor: University of Michigan Press.

Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review, 69,* 109-135.

Feller, W. (1950). *An introduction to probability theory and its applications.* New York: Wiley.

Hays, W. L. (1973). *Statistics for the social sciences.* New York: Holt, Rinehart & Winston.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 276-287.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39,* 341-350.

Karmarkar, U. S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance, 21,* 61-72.

Keeney, R. L. (1968). Quasi-separable utility functions. *Naval Research Logistics Quarterly, 15,* 551-565.

Keeney, R. L. (1971). Utility independence and preferences for multiattributed consequences. *Operations Research, 19,* 875-893.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs.* New York: Wiley.

Krantz, D. H. (1974). Measurement theory and qualitative laws in psychophysics. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Measurement, psychophysics, and neural information processing* (Vol. 2, pp. 160-199). San Francisco: W. H. Freeman.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.

Luce, R. D., & Narens, L. (1985). Classification of concatenation measurement structures according to scale type. *Journal of Math-*

ematical Psychology, 29, 1–72.

Luce, R. D., Raiffa, H. (1957). *Games and decisions: Introduction and critical survey.* New York: Wiley.

Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3, pp. 249–410). New York: Wiley.

MacCrimmon, K. R., & Larsson, S. (1979). Utility theory: Axioms versus paradoxes. In M. Allais & O. Hagen (Eds.), *Expected utility hypotheses and the Allais paradox* (pp. 333–409). Dordrecht, The Netherlands: Reidel.

McNeil, B. J., & Pauker, S. G. (1982). Optimizing patient and societal decision making by the incorporation of individual values. In R. L. Kane & R. A. Kane, (Eds.), *Values and long-term care* (pp. 215–230). Lexington, MA: D. C. Heath.

McNeil, B. J., Weichselbaum, R., & Pauker, S. G. (1978). Fallacy of the five-year survival rate in lung cancer. *New England Journal of Medicine, 299,* 1397–1401.

McNeil, B. J., Weichselbaum, R., & Pauker, S. G. (1981). Speech and survival: Tradeoffs between quality and quantity of life in laryngeal cancer. *New England Journal of Medicine, 305,* 982–987.

Miyamoto, J. M. (1983). Measurement foundations for multiattribute psychophysical theories based on first order polynomials. *Journal of Mathematical Psychology, 27,* 152–182.

Miyamoto, J. M. (1985). *The utility of survival duration and health quality: A conjoint measurement analysis* (Publication No. 8512473). Ann Arbor, MI: University Microfilms.

Miyamoto, J. M. (1987). Constraints on the representation of gambles in prospect theory. *Journal of Mathematical Psychology, 31,* 410–418.

Miyamoto, J. M. (in press). Generic utility theory: Measurement foundations and applications in multiattribute utility theory. *Journal of Mathematical Psychology.*

Miyamoto, J. M., & Eraker, S. A. (1985). Parameter estimates for a QALY utility model. *Medical Decision Making, 5,* 191–213.

Narens, L., & Luce, R. D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin, 99,* 166–180.

Pauker, S. G. (1976). Coronary artery surgery: The use of decision analysis. *Annals of Internal Medicine, 85,* 8–18.

Pauker, S. G., & McNeill, B. J. (1981). Impact of patient preferences on the selection of therapy. *Journal of Chronic Disease, 34,* 77–86.

Pliskin, J. S., Shepard, D. S., & Weinstein, M. C. (1980). Utility functions for life years and health status. *Operations Research, 28,* 206–224.

Raiffa, H. (1969). *Preferences for multi-attributed alternatives* (RM-5868-DOT/RC). Santa Monica, CA: The Rand Corporation.

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Scheffé, H. (1959). *The analysis of variance.* New York: Wiley.

Schoemaker, P. J. H. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature, 20,* 529–563.

Torrance, G. W., Thomas, W. H., & Sackett, D. L. (1972). A utility maximization model for evaluation of health care programs. *Health Services Research, 7,* 118–133.

Tversky, A. (1967). Additivity, utility and subjective probability. *Journal of Mathematical Psychology, 4,* 175–202.

Tversky, A. (1969). Intransitivity of preference. *Psychological Review, 76,* 31–48.

Tversky, A. (1975). A critique of expected utility theory: Descriptive and normative considerations. *Erkenntnis, 9,* 163–173.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Weinstein, M. C., Fineberg, H. V., Elstein, A. S., Frazier, H. S., Neuhauser, D., Neutra, R. R., & McNeil, B. J. (1980). *Clinical decision analysis.* Philadelphia, PA: W. B. Saunders.

Weinstein, M. C., & Stason, W. B. (1982). Cost-effectiveness of coronary artery bypass surgery. *Circulation, 66,* 56–66.

Appendix

Formal Derivations

Proof 1. To prove that the multiplicative utility model predicts that utility independence is satisfied, assume that the multiplicative utility model and Postulates 1 and 2 are valid. We must show that Y'' is the certainty match of $(Y, .5, Y')$ in health state Q if and only if it is the certainty match of $(Y, .5, Y')$ in any other health state Q'.

$Y'' \sim_p (Y, .5, Y')$ in health state Q	
iff	
$(Y'', Q) \sim_p [(Y, Q), .5, (Y', Q)]$	by definition of
iff	the notation
$U(Y'', Q) = U[(Y, Q), .5, (Y', Q)]$	by Postulate 1
iff	
$U(Y'', Q) = sU(Y, Q) + tU(Y', Q)$	by Postulate 2
iff	
$F(Y'')G(Q) = sF(Y)G(Q) + tfF(Y')G(Q)$	assuming the
iff	multiplicative model
$F(Y'')G(Q') = sF(Y)G(Q') + tF(Y')G(Q')$	replacing $G(Q)$
iff	by $G(Q')$
$U(Y'', Q') = sU(Y, Q') + tU(Y', Q')$	assuming the
iff	multiplicative model
$U(Y'', Q') = U[Y, Q'), .5, (Y', Q')]$	by Postulate 2
iff	
$(Y'', Q') \sim_p [(Y, Q'), .5, (Y', Q')]$	by Postulate 1
iff	
$Y'' \sim_p (Y, .5, Y')$ in health state Q'	by definition of
	the notation

Therefore, Y'' is the certainty match of $(Y, .5, Y')$ in health state Q if and only if it is the certainty match of $(Y, .5, Y')$ in health state Q'. This completes the proof that the multiplicative utility model predicts that survival duration is utility independent of health quality.

Proof 2. To prove that short-term indifferent subjects must produce lower certainty matches for some gambles when inferior health is assumed, assume that Postulates 1 and 2 hold, and that an individual is short-term indifferent in the sense of Figure 1. We must show that there exist gambles for survival duration that have lower certainty matches when the inferior health state is assumed. Let Y_0 denote the duration at which the Q_1 and Q_2 utility functions begin to diverge. Thus, the functions are identical below Y_0 and separate above Y_0. Choose any Y and Y' such that (a) $Y > Y_0 > Y'$, and (b) the certainty match of $(Y, .5, Y')$ in Q_1 is less than Y_0. It is always possible to find values of Y and Y' that satisfy these conditions by bringing Y' closer to zero or Y closer to Y_0. Let Y'' denote the certainty match of $(Y, .5, Y')$ in Q_1 and let Z denote the certainty match in Q_2. We need to show that $Y'' > Z$. Because $Y > Y_0 > Y'$, $U(Y, Q_1) > U(Y, Q_2)$ and $U(Y', Q_1) = U(Y', Q_2)$. Therefore $U(Y'', Q_1) = sU(Y, Q_1) + tU(Y', Q_1) > sU(Y, Q_2) + tU(Y', Q_2) = U(Z, Q_2)$. But $Y'' < Y_0$, so $U(Y'', Q_2) = U(Y'', Q_1) > U(Z, Q_2)$. Hence, $Y'' > Z$, which was to be proven.

Received September 10, 1986
Revision received August 4, 1987
Accepted September 21, 1987 ■