The Standard Gamble Method: What Is Being Measured and How It Is Interpreted

Amiram Gafni

Objective. The Standard Gamble (SG) technique is recommended for measurement of individuals' preferences under uncertainty and to express the outcome of different therapeutic choices in utility values to be used in clinical decision analysis and health program evaluation. The article alerts users of this technique to problems stemming from inappropriate interpretation of results of measurements using the SG method.

Study Design. We review different situations where the SG method is used to measure individuals' or group preferences.

Principal Findings. We demonstrate inappropriate interpretation of the time dimension at the individual level; issues stemming from the aggregation of individual utility values measured using different time horizons; the potential for double counting of the time preference effect when discounting future quality-adjusted life years (QALYs); and problems associated with using the SG technique to measure temporary health states.

Conclusions. The inappropriate interpretations stem mainly either from ignoring the time dimension, which is inextricably bound to the health of the individual, or from adding assumptions, in addition to those required by von Neumann-Morgenstern (vNM) utility theory, that are not supported by empirical evidence. An alternative approach to QALYs, the healthy years equivalent (HYE), which incorporates the SG but avoids many of these problems, is described.

Keywords. Standard gamble, utility theory, QALYs, HYEs

When comparing health care programs, two of the most important outcome attributes are survival duration and quality of life. For purposes of program evaluation and clinical decision analysis, it is frequently desirable to measure health outcomes as a single score that takes into account trade-offs between quality of life and survival durations. Following the axiom that individuals are the best judges of their own welfare, this score should also stem from individuals' preferences since ignoring these preferences in the process of decision making can result in choosing the wrong (i.e., less preferred) service (see, for example, McNeil, Weichselbaum, and Pauker 1978; Ben Zion and Gafni 1983; and Mehrez and Gafni 1987).

One way to fully incorporate patients' (or individuals') preferences is to use von Neumann–Morgenstern (vNM) utility theory to generate forms of utility functions over life years and health states to be used in evaluating the outcomes of different interventions (see, for example, Pliskin, Shepard, and Weinstein 1980; and Miyamoto and Eraker 1985, 1988, 1989). Indeed, vNM utility theory is seen by many health services researchers as the "gold standard" for modeling rational behavior under uncertainty both at the individual level (e.g., Pauker and Kassirer 1987) and the group level (e.g., Torrance 1986). The theory tells us how individuals should make decisions under conditions of uncertainty based on a set of behavioral axioms (e.g., Farquhar 1984), and in the process, how to measure individual preferences (i.e., utilities, which are the numbers that represent the strength of the individual's preference for particular outcomes when faced with uncertainty). However, expressing the outcomes of different therapeutic choices in utility values can create problems of interpretation. For example, the meaning of cost per util gained may not be clear or intuitively appealing to many users.

The most commonly used, and intuitively appealing, measure of outcome in cases where trade-offs exist between quality of life and years of survival is the quality-adjusted life-years (QALYs) gained. In brief, the QALY approach is essentially a weighting scheme where the time spent in ill health (measured in years) is multiplied by a weight measuring the relative desirability of the illness state to yield a number that represents the equivalent number of years with full health. Like utilities, QALYs provide a common unit of measure that allows comparisons across programs. Unlike utilities, the calculation of QALYs only partially incorporates individuals' preferences, since the utility approach is used *only* to obtain the weights that go into the QALY calculations (see Mehrez and Gafni 1989). Further, it is known that only under very restrictive conditions is a utility-weighted

---

This work was supported by an educational grant from Sterling-Winthrop Faculty of Economics. The views expressed in this work are the views of the author and do not necessarily reflect the views of the sponsor.

Address correspondence and requests for reprints to A. Gafni, Ph.D., Professor, Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada L8N 3Z5. This article, submitted to *Health Services Research* on April 3, 1992, was revised and accepted for publication on July 30, 1993.

QALY a utility (Mehrez and Gafni 1991). As already stated by proponents of the utility-based QALY, "These conditions... are uncommon in practice, and thus generally a utility-weighted QALY is not in itself a utility" (Torrance and Feeny 1989, 569). However, when the weights are calculated using the Standard Gamble approach (i.e., a lottery-based technique), it is seen by many as an advantage because it associates the index with the rigorous theoretical foundations of the vNM utility theory (see Torrance 1986; Torrance and Feeny 1989; Froberg and Kane 1989).

The Standard Gamble (SG) technique is a classic method of measuring preferences under uncertainty. It is used to measure vNM utility functions over life-years and health states as well as the preference weights to be used in the QALY calculations. The purpose of this article is to alert users of this technique to different problems stemming from inappropriate interpretation of the SG measurement results.

WHAT DO WE MEASURE WHEN USING THE SG TECHNIQUE?

The Standard Gamble (SG) is based directly on the axioms of vNM utility theory (or expected utility theory). For a review of this method and its application in general, see Farquhar (1984) and for its use in health care see Torrance (1986). For simplicity, but without loss of generality, most of this article deals only with the case of estimating a utility value for a chronic health state. In this situation an individual has a particular number of remaining life-years in a given constant health state (e.g., end-stage renal disease) followed by death.

We use the following notations and definitions: Let Q and T denote two attributes of the outcome of concern (Q = health state of the individual, T = remaining life-years). Let $\overline{Q}$ represent the state of full health and $\underline{Q}$ death ($\underline{Q} \leq Q \leq \overline{Q}$). Let $U(Q,T)$ be a von Neumann–Morgenstern utility function that describes the utility of being in a given health state Q, starting now, for a period of T years, followed by death, as viewed now by the individual. For the case of chronic health states (that are preferred to death), the SG technique is applied as follows (see also Figure 1):

> The subject is offered two alternatives. Alternative 1 is a treatment with two possible outcomes: either the patient is returned to normal health and lives for an additional T years (probability p), or the patient dies immediately (probability $1 - p$). Alternative 2 has the certain outcome of chronic state i for life (T years). Probability p is varied until the respondent is indifferent between the two alternatives at which point the required preference value for state i (Q using the above notation) is simply p^*. (Torrance 1986, 20)

Using the notation defined above it can be shown that $U(Q,T)$, the preference value of living T years in health state Q, is equal to p^*.[1]

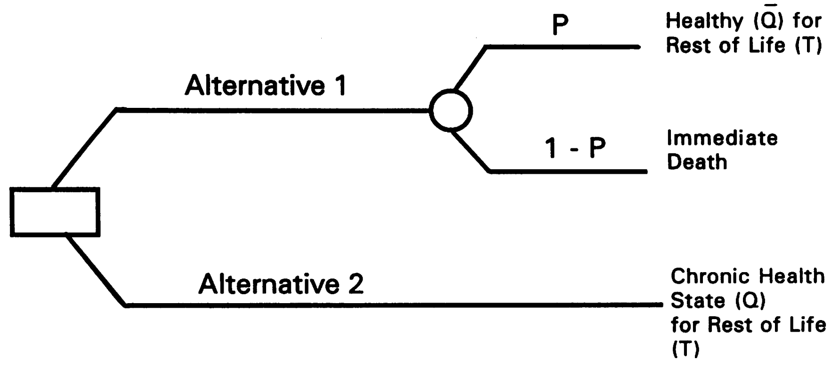
The health of an individual is unlike many other outcomes studied in decision theory or economics. For example, the health of an individual has a time aspect *inextricably bound to it.* Thus, one cannot measure only the preference value attributed by individuals to different health states *ignoring the time spent in this health state.* In this sense, health is a two-dimensional phenomenon. In particular, Gafni and Torrance (1984) argue that the utility for additional time in a given health state–measured using the SG method–depends on a quantity effect (related to the duration in the state), time effect (reflecting time preference), and gambling effect (reflecting risk attitude). Thus, observed values of $U(Q,T)$ already embody the individual's risk attitude, time preference pattern, and attitude toward additional quantities of life.

In the next sections we discuss problems stemming from inappropriate interpretation of utility measurements derived from the SG technique. These problems are discussed in the context of either utility or QALY calculations.

INAPPROPRIATE INTERPRETATIONS OF THE TIME DIMENSION

The SG technique provides an individual's preference score for living in a given health state for a given period of time, that is, $U(Q,T)$. Yet many researchers interpret this value as a "timeless" one. They define a general utility scale where a score of 1.0 represents a normal or healthy health state and 0.0 represents the health state dead (see, for example, Torrance 1986,

Figure 1: The Standard Gamble Method for Eliciting Utilities for a Chronic Health State (Q) for Rest of Life (T)



1987; Torrance and Feeny 1989; Goel, Deber, and Detsky 1989; Boyd, Sutherland, Heasman, 1990; and Weeks, Tierney, and Weinstein 1991). Interpreting the values measured as "timeless" allows the development of tables (e.g., see Torrance and Feeny 1989), in which different health states are organized in declining order from healthy (1.0) to death (0.0), regardless of the time spent in each health state.

Another way to interpret these values is assuming that a "constant proportional trade-off" exists between quality of life (Q) and time (T), which has been shown to be a prerequisite for using SG weights in QALYs (see Pliskin, Shepard, and Weinstein 1980). Under this assumption, the proportion of remaining life that one would trade for a specific quality improvement is independent of the amount of remaining life. If this assumption holds, one can construct a general utility scale where 1.0 represents being healthy for T years, 0.0 represents immediate death, and any value in between represents being in a given intermediate health state for T years. Given the constant proportional trade-off assumption, the preference value of the intermediate health state will be the same regardless of T (the time spent in the state); hence, this value can be interpreted as the value of the health state (similar to the "timeless" interpretation).[2]

It is important to emphasize that the constant proportional trade-off assumption is in addition to assumptions underlying the theoretical foundations of expected utility theory (or vNM utility theory). In other words, an individual can be an expected utility maximizer without following this particular assumption. Indeed, many who have invoked this assumption admit that it is very restrictive and unlikely to represent individuals' behavior (see, for example, Pliskin, Shepard and Weinstein 1980; Miyamoto and Eraker 1985; Torrance and Feeny 1989). Further, a review of the literature by Loomes and McKenzie (1989) has concluded that this assumption (as well as others needed to equate QALY to a utility) is not supported by empirical evidence.

In treating health state utilities as "timeless," we are thus at risk of misrepresenting the actual preferences of people. The following example illustrates the point. Suppose an individual is asked to evaluate the following two scenarios: one being in chronic renal failure treated in a dialysis unit for ten years followed by death, the other being in chronic renal failure treated by a kidney transplant followed by ten years survival. Using the SG approach we find that the utility value of the former is, say, 0.6 and the latter is, say, 0.7. In other words the individual prefers the kidney transplant to the dialysis unit scenario if the survival in both is ten years. Assume now that the individual is being asked to evaluate the same two treatment scenarios, but in this case survival is assumed to be six months followed by death in each scenario. If the constant proportional trade-off assumption holds, the corresponding utility values stay the same as those for ten years. However, this assumption may not hold (i.e., individuals' response to the SG questions might not provide identical utility scores). Using the SG questions we may find that the utility value of the former is now 0.75 and the latter, 0.65. In other words, the individual might prefer the dialysis unit scenario to the kidney transplant scenario if the survival under both is only six months. Such behavior is not unrealistic. For example, Sutherland et al. (1982) examined individual preferences for various health states (including death) under certainty and for various time periods. They found that for a short period of survival (three months) the majority of their subjects preferred all health states to death. However, for a longer survival (eight years) there is a reversal of preference for the lower-ranked health states: the majority of subjects preferred death.

It is also important to emphasize that these are all legitimate responses for an individual who follows expected utility theory. There is nothing irrational in the individual's responses in this example. Only if we adopt the assumption about constant proportional trade-off are we able to predict the utility scores for the second set of questions from the scores derived from answers to the first set. As can be seen the "timeless" utility (or constant proportional trade-off) assumption greatly simplifies the application of multiperiod utility models to health care program evaluations because it supports the use of the standard QALY formula. However, the 'price' paid is that the scores derived under this assumption may not reveal the true preferences of individuals; in the example just cited, the 'patient's' preferences were reversed.

Use of "timeless" preference scores may also lead to double counting of the time dimension in the process of calculating QALYs. Because the QALY approach is the summation of years of life adjusted for quality, the time dimension appears twice: first in the number of years spent in a given health state and second in the weight used for the quality adjustment (i.e., $U(Q,T)$ when using the SG method, as described earlier). Using the notation described earlier, the QALY calculation involves the following multiplication: $U(Q,T) \cdot T$.[3] However, researchers often interpret the expression as if it were $U(Q) \cdot T$. This interpretation reflects the tendency to use data from the published literature irrespective of the time dimensions in the study where the scores were originally derived and the one in which they are going to be used. This again can create distortions of individuals' true preferences (see examples in Mehrez and Gafni 1989).

An alternative measure of outcome, the healthy-years equivalents (HYE), was suggested by Mehrez and Gafni (1989). This measure, based on the theoretical foundations of utility theory, stems directly from the individual's utility function, thus fully reflecting his or her preferences. It combines outcomes of both quality of life (morbidity) and survival (mortality) and thus can serve as a common unit of measure for all programs, allowing comparisons across programs. It preserves the appealing intuitive meaning of the QALY but uses a holistic approach (i.e., describing a life-time health profile). For individuals who are expected utility maximizers (i.e., subscribe to the axioms of vNM utility theory), a simple two-stage, lottery-based (i.e., SG questions) procedure is described in Mehrez and Gafni (1991). This results in outcome measures that do not suffer from the cited problems.

Using the definitions and notations of the previous sections, the HYE approach is as follows: Assume an individual faces a potential outcome of living the rest of his life (T years) in health state Q. Let H^* be the healthy-years equivalent of (Q,T), defined as

$$U(\overline{Q},H^*) = U(Q,T) \tag{1}$$

Thus, the individual is indifferent between H^* years in full health ($\overline{Q}$) and T years in Q.

For a vNM individual the HYE is measured as follows. In the first stage the individual is presented with a lottery with two possible outcomes: full health for T years (probability p) or immediate death (probability $1 - p$). The second alternative offers the outcome of living in the chronic state (Q) for the remaining years (T). Probability p is varied until the individual is indifferent between the alternatives. Denoting the utility of full health for T years as 1.0 and the utility of death as 0.0, at the indifference point the individual's utility of living T years in health state Q is equal to p^* (the probability at the indifference point), i.e., $U(Q,T) = p^*$. Note that this stage is identical to the SG question described earlier, which is also used to measure the weights in the QALY calculations (assuming those weights are derived using the SG method).

In the second stage, lottery questions are used to convert the time in ill health (i.e., the chronic state) to time in full health (HYEs). The individual is offered two alternatives. The first is a lottery with two possible outcomes: full health for T years (probability p^*) or immediate death (probability $1 - p^*$). Note that p^* is the value of the probability at the indifferent point in stage one. The second alternative is the outcome of living H more years in full health. H is varied until the person is indifferent between the alternatives. H^* denotes the value of H at the indifference point and represents the hypothetical number of years in full health, which is equivalent from the individual's perspective to living T years in health state Q. Note that in this procedure we do not make any assumptions about the form of the individual's utility function. On the contrary, the individual's own responses (i.e., stated preferences) are used to convert the years in ill health to years in full health.

AGGREGATION OF INDIVIDUAL UTILITIES MEASURED USING DIFFERENT TIME HORIZONS

In cost-utility analysis "the aggregation across subjects is achieved by measuring all individual utilities on the common 0–1 dead–healthy scale and taking the arithmetic mean" (Torrance 1986, 17). This simple (nonweighted) mean of individuals' responses is used to derive a "social valuation" of the weights (per unit of time) to be used in QALYs calculations. The question of aggregation of individual utilities and its validity has been addressed by many authors (e.g., Arrow 1963; Harsanyi 1955, 1975; Keeney and Raiffa 1976; and Kalai and Schmeidler 1977, to name a few). In this section we do not deal with the question whether or not it is valid to aggregate individual utilities but *assume* that it can be done (albeit under a very restrictive set of assumptions).

In this section we question the meaning of aggregating individual utility scores measured using different time horizons, based on the simple arithmetic mean. As explained in the previous section, each utility score on the scale healthy-to-dead represents the utility of living a period of time in a given health state (i.e., $U(Q,T)$. For the chronic health state (which we use as an example in this article), the time span is usually defined as the individual's expected life span. In many cases individuals who participate in a study have different expected life spans. Thus the simple aggregation rule implies subscribing to the equity criterion that the reference state of full health for the rest of the individual's life should be treated as of equal value for all individuals. But the equity criteria used to justify this method of aggregation differ from the one indicated above. For example, Torrance (1986) adopts as the appropriate equity criterion that "the difference in utility between being dead and being healthy is set equal across people" (p. 17), while in later work Feeny and Torrance (1989) adopt as an appropriate basis for aggregation that "the utility of full healthy life from birth is set equal for each individual" (p. 193). As noted by Gafni and Birch (1991), these criteria do not coincide, and neither is consistent with the simple aggregation rule recommended and used in the literature.

Whether a particular equity criterion is appropriate or not is a subjective issue. However equity considerations (i.e., the relative values or weights attributed to different individuals or groups) are an intrinsic part of any evaluation. It is thus important that consideration is given to the question of whether the procedure measuring the outcome is consistent with the stated equity criteria. Gafni and Birch (1991) deal with this issue extensively and derive adjustment algorithms based on the axioms of vNM utility theory and taking into account the different equity criteria adopted in the literature.

This discrepancy, described by Gafni and Birch (1991), between the equity statement and the measurement procedure holds regardless of whether we accept or reject the assumption about constant proportional time trade-off. Under this assumption the interpretation of the utility values is this: the utility of being in health state Q for rest of life (T) is equal to α ($0 < \alpha < 1$) multiplied by the utility of being in full health for T years. The hypothesis to be tested is that α is equal for all individuals (i.e., regardless of T). However, if we do not accept this assumption (due to the lack of empirical evidence to support it), it casts further doubt on the meaning of the simple arithmetic mean. The following example illustrates the point.

Assume a group that consists of two individuals. Both are in health state Q. However, individual 1 has 10 years to live and individual 2 has 20 years to live in this health state. Using the SG technique to measure $U(Q,10)$, we offer individual 1 two alternatives: a lottery with two possible outcomes (living 10 years in full health followed by death, or immediate death) versus a certain outcome of living 10 years in health state Q. To measure $U(Q,20)$, individual 2 is again offered two alternatives: a lottery with two possible outcomes (living 20 years in full health followed by death, or immediate death) versus a certain outcome of living 20 years in health state Q.

Note that in spite of the fact that the health state is identical in both cases, the time spent in this health state is different (10 years for individual 1 and 20 years for individual 2). Consequently, different evaluation spaces were used to measure individuals' preferences for the outcomes. In the case of individual 1, the upper bound (or best-case scenario) of the evaluation space is living 10 years in full health; in the case of individual 2 it is living 20 years in full health. The upper bound of the evaluation space is important since it is used as a reference state and is assigned an arbitrary value of 1.0. But only under the constant proportional trade-off assumption can the utility scores be interpreted as a "period utility" (i.e., each utility value represents a ratio of the utility of T years at state Q to the utility of T years at state full health). This means that in our example, the proposed aggregation procedure implies that living 20 years in full health is equal to living 10 years in full health. In other words, living 10 additional years in full health has a utility of zero.[4]

Finally, we can use the above example to illustrate the effect of the choice of an equity criterion on the value of the group utility. Assume that we apply the measurement method and the aggregation rule of an arithmetic mean to the above example. Assume that the response of individual 1 to the SG resulted in a utility score of $U(Q,10) = 0.7$ (i.e., living 10 years in health state Q has a utility value of 0.7) and the response of individual 2 resulted in a utility value of $U(Q,20) = 0.6$. Using a simple arithmetic mean the group utility is equal to 0.65. But this is based on the equity criterion that the reference state of full health for the rest of the individual's life should be treated as of equal value for all individuals (i.e., $U(\textit{Full health}, 10\ \textit{years}) = U(\textit{Full health}, 20\ \textit{years}) = 1.0$).

Let us now use a different equity criterion, for example, that the utility of full healthy life from birth is set equal for each individual (Feeny and Torrance 1989). This implies that $U(\textit{Full health}, \textit{Full life span}) = 1.0$, which means that $U(\textit{Full health}, 10\ \textit{years}) < U(\textit{Full health}, 20\ \textit{years}) < 1.0$. Using the adjustment algorithm described in Gafni and Birch (1991), suppose we find that $U(\textit{Full health}, 10\ \textit{years}) = 0.15$ and $U(\textit{Full health}, 20\ \textit{years}) = 0.4$. This implies that $U(Q, 10) = 0.105\ (0.7 \cdot 0.15)$ and $U(Q, 20) = 0.24$, and the group utility now equals 0.1725. It is easy to see that choosing the second equity criterion results in a different set of group scores. Note also that on the new scale (death = 0.0; full health from birth to death = 1.0) we find that $U(Q, 20) > U(Q, 10)$. However, when comparing the numbers on the old scale (death = 0.0; full health for rest of life = 1.0), they have meaning only under the strong assumption of constant proportional trade-off over time.

DISCOUNTING OF FUTURE QALYS GAINED

Effects occurring in the future require discounting in order to incorporate individuals' time preferences (see, e.g., Drummond, Stoddart, and Torrance 1987; Weinstein, Fineberg, Elstein, et al. 1980; Shepard and Thompson 1979; Weinstein and Stason 1977; Levine 1975; Torrance, Sackett, and Thomas 1973; Bush, Chen, and Patrick 1973). Most researchers therefore discount future gains measured in QALYs (Drummond, Stoddard, and Torrance 1987; Weinstein and Stason 1977; Torrance, Sackett, and Thomas 1973). In this section we do not discuss the question of whether and how to discount the future health benefits of competing public programs (this is dealt with in Lipscomb 1989, and in Krahn and Gafni 1993); rather, we alert readers who follow the current recommended methodology that inappropriate interpretation of the time dimension usually results in double counting of time preferences. As in the previous sections we demonstrate our argument using the case of a chronic health state.

As recommended (e.g., Drummond, Stoddard, and Torrance 1987), the QALY estimate is the product of the utility score for the scenario measured multiplied by the discounted time spent in the scenario. Using the notations and definitions from previous sections and denoting D_T as the appropriate discount factor,

$$D_T = \sum_{i=1}^{T} = 1/(1+r)^i \tag{2}$$

where r is the discount rate, the QALY value of living T years in health state Q followed by death is equal to $U(Q,T) \cdot T \cdot D_T$. Note that $U(Q,T)$ is the preference score attributed by the individual to living T more years in health state Q followed by death. This score was measured in this example using the SG technique described earlier.

Gafni and Torrance (1984) argue that responses to hypothetical health lottery questions, such as those used to measure $U(Q,T)$, already capture the time preference pattern of the individual. For example, if the individual has a positive time preference (prefers that good things will occur sooner), this preference is already taken into account in his or her answers to the lottery questions used to measure $U(Q,T)$. Because $U(Q,T)$ is used as the weight in the QALY calculation, using an additional discount factor (D_T) simply results in a double counting of the time preference effect and might affect the result of an analysis.

Eliminating "double discounting" is not an easy task. The only attempt we know of to deal with this issue is in a recent article by Lipscomb (1989). Following Gafni and Torrance (1984), the article acknowledges that "risk attitude," "time preferences," and "quantity effects" are among the factors that affect the overall choice of an individual's therapy. The work describes a practical, two-step evaluation procedure featuring the "scenario strategy," a holistic multiattribute preference approach to evaluating multiperiod health outcomes. The method suggested in the article allows one to isolate statistical time preference effects at the individual or group level (i.e., to test for the existence and strength of a distinct time preference effect and determine the functional form of the discount rate [e.g., constant, time dependent]). Thus, this method has the potential to help researchers correct for double counting of the time effect if it exists.

More specifically, the scenario strategy involves: *Step 1*–Using a factorial design, a large set of multiperiod, multistate health outcome scenarios is specified. *Step 2*–Each individual in the sample assigns a relative preference score to each scenario using a variety of competing techniques. *Step 3*–Multivariate regression analysis is used to explain variations in the preference scores. *Step 4*–The estimated model (*Step 3*) is used to infer time preference, quantity gambling, and possibly other effects. For a more detailed description of this method and examples of application, see Lipscomb (1989).

TEMPORARY HEALTH STATES AND THE SG METHOD

The SG method is also recommended as a way of measuring individuals' preferences for temporary health states:

> Here intermediate states i are measured relative to the best state (healthy) and the worst state (temporary state j).... In the basic format... the formula is $h_i = p + (1 - p)h_j$, where i is the state being measured and j is the worst state. If death is not a consideration in the use of utilities, h_j can be set equal to zero and the h_i values determined from the formula which then reduces to $h_i = p$. (Torrance 1986, 21–23)

This description of how to measure preferences toward temporary health states implies that the utilities measured are "timeless." However, as mentioned before, the health of an individual has a time aspect inextricably bound to it, and one cannot measure a preference value of a health state while ignoring the time spent in this health state. Furthermore, temporary health states lead to other health states (by definition) which are not always death.

The case of temporary health states leads to a broader methodological issue, that is, whether one can measure the preference value for a given health state for a standardized unit of time (say a year; see, e.g., Weeks, Tierney, and Weinstein 1991). In this case, using the SG technique the individual will be offered two alternatives: a lottery with two possible outcomes (perfect health for one year or death) versus a certain outcome of being in health state Q for one year. Theoretically, this is a viable alternative; in practice, it might be problematic. The main difficulty stems from the need to incorporate in the description of the health state an explicit statement about the future of the individual after the year ends. If this is not done (e.g., Weeks, Tierney, and Weinstein 1991) one does not know what the respondent assumed about the future health state (beyond one year) when responding to the SG question. If individuals assume different post-one-year future states, their responses will embody an additional component of variation that cannot be teased out.

Two options are available. The first approach is to ask the individual to imagine living only one year, followed by death, in each of the outcome alternatives. This is likely to be seen as threatening by participants, especially when they are evaluating less severe health states (e.g., side-effects of hypertension) and are not *very* old, so that many more years of life are anticipated. The second alternative is to ask the individual to imagine that after the year is over, he or she will live the rest of life in a state of full health. This is less threatening, but it means that we now measure the individual's preference value for the profile "living one year in a given health state followed by many more years in full health." Again, this preference value is different from the one which QALY proponents would like to measure–that is, for living one year in the given health state. Strictly speaking, to properly assess the utility of living one year in a given health state one has to know the specific functional form of the individual's lifetime utility function. This is not an easy task (to say the least) and may not be feasible in the context of an empirical study.

Measuring a temporary health state in isolation (i.e., ignoring other health states that follow or precede the one measured) is methodologically problematic. A recent study (Hall et al. 1992) that evaluated the cost utility of mammography screening in Australia used the TTO[5] technique to measure the outcome. The authors report that their

> data allowed for testing of the hypothesis that the value of the combined scenario could be derived from the independently assessed values of the three health states. The results indicate that the utility associated with a complex scenario may not be accurately calculated by a weighted average of the utilities of the constituent states and the assumption of a reasonable time preference. It appears that the prognosis has a significant effect upon the assessment of the preceding health states and therefore the holistic approach must be adopted to obtain valid results (p. 996).

Although the method used to measure preferences in the Hall et al. (1992) study was not the SG approach, there is no a priori reason why similar results would not be likely using the SG method. For the results to differ, we have to assume that individuals' utility functions under uncertainty are separable and additive over time (Mehrez and Gafni 1991; Broome 1993). Intertemporal additivity or strong separability has never been viewed as normatively compelling since there are many situations (health included) in which it is reasonable for consumption at one point in time to influence the marginal utility of consumption at another (Loewenstein and Prelec 1991). As Broome states, "The most dubious condition is that, in the person's preferences, qualities of life at different times are strongly separable . . . strong separability means that a person's preferences about the qualities of her life in any particular group of years are independent of the qualities of her life in other years" (p. 152). Thus, we support the suggestion by Hall et al. (1992) that the holistic approach should be adopted to obtain valid results, and Mehrez and Gafni (1991) describe algorithms to measure HYEs using the holistic approach.

CONCLUSION AND DISCUSSION

Von Neumann-Morgenstern utilities are seen by many as an important tool for clinical decision analysis under uncertainty regarding alternative treatments for individual patients. They are also often aggregated across individuals to provide group utility scores. These group utility scores are used as weights in the process of calculating QALYs for use in program evaluations. In this article we alert users of the SG method to different problems stemming from inappropriate interpretation of the SG measurement results. Those inappropriate interpretations stem mainly from either ignoring the time dimension, which is inextricably bounded to the health of the individual, or from adding assumptions (i.e., in addition to those required by expected utility theory). An important one is that of constant proportional trade-off, which is not supported by existing empirical evidence (Loomes and McKenzie 1989). Although this is a crucial assumption, we are not aware of attempts to validate it even in studies that rely on it to interpret results.

A major source of confusion is researchers' search for a "timeless" weight to be used in the process of calculating QALYs. The use of utilities, measured by SG method, is particularly appealing to many since it apparently associates the index (QALYs) to a rigorous theoretical foundation. As we have shown in this work, the use of the SG method for this purpose is problematic. Indeed, it is impossible to measure utility values for timeless health states when using the SG technique. Mehrez and Gafni (1989) have suggested a different measure of outcome–healthy-years equivalent (HYE), which stems directly from the individual's utility function and thus better represents the individual's preferences. This measure can be used for both clinical decision making and program evaluation and, like QALYs, it is intuitively appealing. Algorithms that describe the process of measuring HYE using the SG approach can be found in Mehrez and Gafni (1991). These algorithms acknowledge the time aspect in individuals' health and thus avoid many of the problems mentioned in this study.

The SG method is also being used in the context of measuring multiattribute utility functions. Lottery type questions, like those described herein, are used to measure the utility scores for the different levels of the different attributes and to help calculate the different parameters of the utility function (e.g., the weights in an additive function). When multiattribute utility functions are used to generate the weights for the QALYs calculations, the analysis suffers from the same problems of interpretation described in this work. Moreover, the multiattribute utility approach requires the assumption about independence of health states over time (i.e., intertemporal additivity) in order to arrive at the value of a combined scenario (e.g., Torrance, Boyle, and Horwood 1982; Drummond, Stoddart, and Torrance 1987). The main advantage of the multiattribute utility approach is that it provides a method to specify the utility scores of a large number of health states from a limited number of utility measurements.

We chose to interpret the results of the use of the SG method in the context of a von Neumann-Morgenstern utility theory because it is seen by many in health as the "gold standard" of rational behavior under uncertainty both at the individual (e.g., Pauker and Kassirer 1987) and group (e.g., Torrance 1986) levels. It is well documented that reasonable people often violate the assumption of expected utility theory, which is a criticism of how well this theory describes individuals' behavior under uncertainty. Numerous empirical studies have commonly found that individuals do not behave consistently with the expected utility theory axioms: when asked to reconcile their actual behavior with the axioms, many people, while accepting the appeal of these axioms, preferred to act in their original fashion. It also must be acknowledged that many leading researchers in the field of decision theory and economics no longer accept this theory as the normative standard (i.e., how individuals should behave) for decision making under uncertainty. Thus, new theories have been designed to accommodate predictable violations of the traditional expected utility model. (For more information see Bell and Farquhar 1986; Fishburn 1988; or Machina 1987). The purpose of this article is *not* to challenge or support vNM utility theory. It is to alert readers to problems that might stem from inappropriate interpretations of the SG results in the context of vNM utility theory when applied to clinical decision making and QALY calculations.

ACKNOWLEDGMENTS

I am grateful for helpful comments from Abraham Mehrez, Bernie O'Brien, and Stephen Birch on an earlier version of this article. However, the responsibility for its contents are solely my own.

NOTES

1. More specifically, using expected utility theory at the indifference point the following relation holds: $U(Q,T) = p^{*}U(\overline{Q},T) + (1 - p^{*})\ U(\underline{Q},T)$ where p^* is the indifference probability. Denoting $U(\overline{Q},T) = 1.0$ and $U(\underline{Q},T) = 0.0$, we have $U(Q,T) = p^{*}$.
2. Note that even if the constant proportional trade-off assumption holds, interpretation of the QALY as a utility score still assumes, *inter alia*, risk neutrality with respect to survival duration in all health states as well as other assumptions.
3. Note that in some applications the two Ts in $U(Q,T) \cdot T$ are not necessarily equal.
4. Since the upper bound is used as reference stat, attributing the value 1.0 to the upper bound in both cases (individuals 1 and 2) implies assuming U(10 *years, full health*) = U (20 *years, full health*) = 1.0.
5. The time trade-off (TTO) technique involves a paired comparison. For example, for the case of chronic health state the comparison is between being in this health

state for the rest of life and being in full health for a shorter period of time. For more on this method see Torrance (1986).

REFERENCES

Arrow, K. J. *Social Choice and Individual Values*. 2d ed. New Haven, CT: Yale University Press, 1963.

Bell, D. E., and P. H. Farquhar. "Perspective on Utility Theory." *Operations Research* 34, no. 1 (1986): 179–83.

Ben Zion, U., and A. Gafni. "Evaluation of Public Investment in Health Care: Is the Risk Irrelevant?" *Journal of Health Economics* 2, no. 3 (1983): 161–65.

Boyd, N. F., H. J. Sutherland, K. Z. Heasman, D. L. Tritchler, and B. J. Cummings. "Whose Utilities for Decision Analysis?" *Medical Decision Making* 10 (1990): 58–67.

Broome, J. "Qalys." *Journal of Public Economics* 50, no. 2 (1993): 149–67.

Bush, J. W., M. M. Chen, and D. L. Patrick. "Health Status Index in Cost-Effectiveness: Analysis of PKU Program." In *Health Status Indexes*. Edited by R. L. Berg. Chicago: Hospital Research and Educational Trust, 1973, pp. 72–208.

Drummond, M. F., G. L. Stoddart, and G. W. Torrance. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford, England: Oxford Medical Publications, 1987.

Farquhar, P. H. "Utility Assessment Methods." *Management Science* 30, no. 11 (1984): 1283–1300.

Feeny, D. H., and G. W. Torrance. Incorporating Utility-Based Quality-of-Life Assessment Measures in Clinical Trials." *Medical Care* 27, no. 3 (Supplement, 1989): S190–S204.

Fishburn, P. C. "Expected Utility: An Anniversary and a New Era." *Journal of Risk and Uncertainty* 1, no. 4 (1988): 267–83.

Froberg, D. G., and R. L. Kane. "Methodology for Measuring Health-State Preferences. II: Scaling Methods." *Journal of Clinical Epidemiology* 42, no. 5 (1989): 459–71.

Gafni, A., and G. W. Torrance. "Risk Attitude and Time Preference in Health." *Management Science* 30, no. 4 (1984): 440–51.

Gafni, A., and S. Birch. "Equity Considerations in Utility-Based Measures of Health Outcomes in Economic Appraisals: An Adjustment Algorithm." *Journal of Health Economics* 10, no. 3 (1991): 329–42.

Goel, V., R. B. Deber, and A. S. Detsky. "Non-ionic Contrast Media: Economic Analysis and Health Policy Development." *Canadian Medical Association Journal* 140, no. 4 (1989): 389–95.

Hall, J., K. Gerard, G. Salkeld, and J. Richardson. "A Cost Utility Analysis of Mammography Screening in Australia." *Social Science and Medicine* 34, no. 9 (1992): 993–1004.

Harsanyi, J. C. "Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63, no. 4 (1955): 309–21.

———. "Nonlinear Social Welfare Functions." *Theory and Decision* 6, no. 3 (1975): 311–32.

Kalai, E., and D. Schmeidler. "Aggregation Procedure for Cardinal Preferences: A Formulation and Proof of Samuelson's Impossibility Conjecture." *Econometrica* 45, no. 6 (1977): 1431–38.

Keeney, R. L., and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. New York: Wiley, 1976.

Krahn, M., and A. Gafni. "Discounting in the Economic Evaluation of Health Care Interventions." *Medical Care* 31, no. 5 (1993): 403–18.

Levine, H. M. "Cost-Effectiveness Analysis in Evaluation Research." In *Handbook of Evaluation Research*, Vol. 2. Edited by M. Guttentag and E. Struening. Beverly Hills, CA: Sage Publications, 1975, 89–122.

Lipscomb, J. "The Preference for Health in Cost-Effectiveness Analysis." *Medical Care* 27, no. 3 (Supplement, 1989): S233–53.

Loewenstein, G., and D. Prelec. "Negative Time Preference." *American Economic Review* 81, no. 2 (1991): 347–52.

Loomes, G., and L. McKenzie. "The Use of QALYs in Health Care Decision Making." *Social Science and Medicine* 28, no. 4 (1989): 299–308.

Machina, M. J. "Choice under Uncertainty: Problems Solved and Unsolved." *Economic Perspectives* 1, no. 1 (1987): 121–54.

McNeil, B. J., R. Weichselbaum, and S. G. Pauker. "Fallacy of Five-Year Survival in Living Cancer." *New England Journal of Medicine* 299, no. 25 (1978): 1397–1401.

Mehrez, A., and A. Gafni. "The Optimal Treatments Strategy–A Patient Perspective." *Management Science* 33, no. 12 (1987): 1602–12.

Mehrez, A., and A. Gafni. "Quality-Adjusted Life-Years, Utility Theory, and Healthy-Years Equivalents". *Medical Decision Making* 9, no. 2 (1989): 142–49.

Mehrez, A., and A. Gafni. "Healthy-Years Equivalents: How to Measure Them Using the Standard Gamble Approach." *Medical Decision Making* 11, no. 2 (1991): 140–46.

Miyamoto, J. M., and S. A. Eraker. "Parameter Estimates for a QALY Utility Model." *Medical Decision Making* 5, no. 2 (1985): 191–213.

Miyamoto, J. M., and S. A. Eraker. "A Multiplicative Model of the Utility of Survival Duration and Health Quality." *Journal of Experimental Psychology* 117, no. 1 (1988): 3–20.

Miyamoto, J. M., and S. A. Eraker. "Parametric Models of the Utility of Survival Duration: Tests of Axioms in a Generic Utility Framework." *Organizational Behaviour and Human Decision Processes* 44, no. 2 (1989): 166–203.

Pauker, S. G., and J. P. Kassirer. "Decision Analysis." *New England Journal of Medicine* 316, no. 5 (1987): 250–58.

Pliskin, J. S., D. S. Shepard, and M. C. Weinstein. "Utility Functions for Life-Years and Health Status." *Operations Research* 28, no. 1 (1980): 206–24.

Shepard, D. S., and M. S. Thompson. "First Principles of Cost-Effectiveness Analysis in Health." *Public Health Reports* 94, no. 6 (1979): 535–43.

Sutherland, H. U., H. Llewellyn-Thomas, N. F. Boyd, and J. E. Till. "Attitudes toward Quality of Survival: The Concept of Maximal Endurable Time." *Medical Decision Making* 2, no. 3 (1982): 299–309.

Torrance, G. W., and D. Feeny. "Utilities and Quality-Adjusted Life-Years." *International Journal of Technology Assessment in Health Care* 5, no. 4 (1989): 559–75.

Torrance, G. W. "Measurement of Health State Utilities for Economic Appraisal: A Review." *Journal of Health Economics* 5, no. 4 (1986): 1–30.

———. "Utility Approach to Measuring Health Related Quality of Life." *Journal of Chronic Disease* 40, no. 6 (1987): 593–600.

Torrance, G. W., M. H. Boyle, and S. P. Horwood. "Application of Multi-Attribute Utility Theory to Measure Social Preferences for Health States." *Operations Research* 30, no. 6 (1982): 1043–69.

Torrance, G. W., D. L. Sackett, and W. H. Thomas. "Utility Maximization Model for Program Evaluation: A Demonstration Application." In *Health Status Indexes*. Edited by R. L. Berg. Chicago: Hospital Research and Educational Trust, 1973, pp. 156–71.

Weeks, J. C., M. R. Tierney, and M. C. Weinstein. "Cost-Effectiveness of Prophylactic Intravenous Immune Globulin in Chronic Lymphocytic Leukemia." *New England Journal of Medicine* 325, no. 2 (1991): 81–86.

Weinstein, M. C., and W. B. Stason. "Foundations of Cost-Effectiveness Analysis for Health and Medical Practices." *New England Journal of Medicine* 296, no. 13 (1977): 716–21.

Weinstein, M. C., H. C. Fineberg, A. S. Elstein, H. S. Frazier, D. Neuhauser, R. R. Neufra, and B. J. McNeil. *Clinical Decision Analysis*. Philadelphia: W. B. Saunders Co., 1980.