Health Status Utility Assessment by Standard Gamble: A Comparison of the Probability Equivalence and the Lottery Equivalence Approaches

Anandi V. Law,[1,4] Dev S. Pathak,[2] and Mark R. McCord[3]

Received July 9, 1997; accepted October 7, 1997

Purpose. Utility values obtained with the standard gamble (SG) method using the probability equivalence approach (PE) have a reported bias due to the "certainty effect." This effect causes individuals to overvalue a positive outcome when it occurs under certainty. Researchers in the decision sciences have proposed an alternative, "lottery equivalence" (LE) approach, using paired gambles, to eliminate this bias. The major objective of the current study was to investigate the certainty effect in health status utility measures and to test our hypothesis that the certainty effect would act in a reverse direction for negatively valued outcomes. ***Methods.*** Fifty-four subjects completed the study by assessing preferences for three health states by rating scale and then by SG using PE as well as LE approaches with assessment lotteries of 0.5 and 0.75. ***Results.*** The results from 41 useable responses point towards possible existence of the certainty effect in health in the hypothesized direction: utility values obtained with the PE were significantly lower than with the LEs. There was no significant difference between the LE values indicating elimination of the bias.
Conclusions. The results have important implications since the SG using PE is thought be the "gold standard" in health status utility measurements.

KEY WORDS: Health status utility assessment; standard gamble approaches; probability equivalence; lottery equivalence; certainty effect.

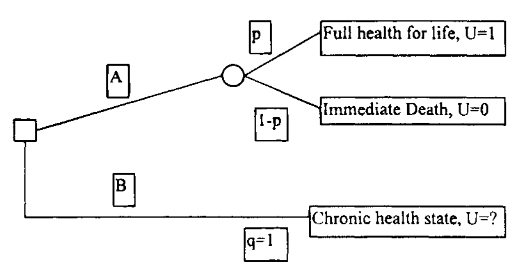
INTRODUCTION

Health state utility can be defined as the preference for, or desirability of, a certain health condition. It has been used to measure the value of health improvement; it is also an underlying measure of health-related quality-of-life (HRQOL) and the outcome measure for cost-utility analysis, both of which are used for clinical decision making and for economic evaluations of health care programs. The utility approach to measuring HRQOL has the advantage of providing a single cardinal unit of HRQOL which is useful in quantitative analyses (1).

Cardinal utilities in health state measurements have been commonly measured via three methods, i.e., the rating scale, the time trade-off and the standard gamble (2). The rating scale consists of a line with anchor points of least and most desirable outcomes and involves placing the health states under consideration between the two anchors, in order of their preference, such that the intervals between placements correspond to the subject's perceived difference in preference (2). The second method, the time trade-off was developed specifically for health care by Torrance et al. and it measures utilities implicitly as opposed to the explicit assessment of preference by the rating scale (2). The subject is offered a trade-off between health state l for time t and full health for time $x<t$ followed by death, and the time x is varied until the subject is indifferent between alternatives. Utility is calculated as $U=x/t$.

The third, and most commonly recommended method is the standard gamble approach. "It is the classical method of measuring cardinal preferences" and incorporates the aspect of risk or uncertainty in decisions (2). It follows the axioms of utility theory as presented by von Neumann and Morgenstern in 1947 (3). In this method, the subject is presented with two alternatives; alternative A is a reference lottery where the best possible outcome of the treatment (that the subject will return to full health and live for an additional t years) occurs with a probability p and the worst possible outcome of immediate death occurs with a complementary probability $1\text{-}p$ (Figure 1). This is countered against alternative B which has a probability q of 1.0 (or certainty) that the chronic health state l will persist for life (t years). The utility of full health (FH) is arbitrarily set at a value of 1.0 and that of immediate death (ID) at 0.0. The probability p is varied until the subject is indifferent between alternatives A and B; this is called the indifference point. The expected utilities of alternatives A and B are equal at this point, i.e.,

$$\begin{aligned} q(l) \times U(l) &= p(FH) \times U(FH) + (1-p)(ID) \times U(ID) \\ \therefore 1 \times U(l) &= p \times 1 + (1-p) \times 0 \qquad (1) \\ \therefore U(l) &= p \end{aligned}$$



At point of indifference,
Expected Utility of A = Expected Utility of B
$\therefore$ [p(FH) × U(FH) + (1-p)(ID) × U(ID)] = q(CHS) × U(CHS)
$\therefore$ p(1) + (1-p) × (0) = 1 × U(CHS)
$\therefore$ U(CHS) = p
where FH=full health, ID=immediate death, CHS=chronic health state
Note: The squares in the decision tree denote a decision node, the circles denote a chance node.

Fig. 1. Standard gamble under probability equivalence.

[1] Division of Pharmacy Practice and Administration, College of Pharmacy, The Ohio State University, Columbus, Ohio 43210.
[2] College of Medicine and Public Health and College of Pharmacy, The Ohio State University, Columbus, Ohio 43210.
[3] Civil and Environmental Engineering and Engineering Graphics, College of Engineering, The Ohio State University, Columbus, Ohio 43210.
[4] To whom correspondence should be addressed. (e-mail: law.2@osu.edu)
An earlier version of this manuscript was presented at the ninth Annual Meeting of the American Association of Pharmaceutical Scientists (AAPS), 8th November, 1995, Miami, Florida.

NOTATION: SG, Standard Gamble; PE, Probability Equivalence; LE, Lottery Equivalence; HRQOL, Health-Related Quality-of-Life; Ss, Subjects; ANOVA, Analysis of variance; α, Alpha or significance level of test.

0724-8741/98/0100-0105$15.00/0

Thus the utility of the assessed health state depends on the indifference probability p of the reference lottery. The method described above is defined in the utility literature as the standard gamble using a probability equivalence method with extreme gambles and has been the method of choice in health status utility measurements (4).

However, the traditional standard gamble under probability equivalence (PE) is susceptible to various problems, namely; biases from range effects due to the extreme values of the anchor points and distortions in risk behavior (4,5). PE is also susceptible to serial dependence especially if responses are chained, i.e., if utility values obtained earlier in a standard gamble are used in later comparisons, any bias in earlier responses is amplified in later ones (5). Further, the major drawback of the PE approach to measuring utility is the presence of a "certainty effect", a pervasive psychological phenomenon whereby an individual values an outcome more highly when it occurs with certainty (5,6). The certainty effect increases with the reference probability levels used in the comparisons, i.e., subjects exhibit greater risk-averseness as the probability in the gamble is increased. The Allais paradox has been cited as a classic example of the certainty effect (6,7). [Please see Appendix 1 for an example of the Allais paradox] (8)

This effect has been discussed in the behavioral decision making literature and documented by McCord & deNeufville in standard gamble measurements using certainty equivalence [a closely related standard gamble method where the individual cites an amount for the certain outcome which makes them indifferent between the lottery and the certainty, and which shares common disadvantages with the PE method] (4,9). McCord and deNeufville found that the certainty effect introduced systematic errors in the standard gamble; subjects in their study overweighted a positively valued outcome (e.g., dollars) since it occurred with certainty and hence the utility of the outcome also tended to be overvalued. They concluded that, for positively valued outcomes, the utility function is shifted above and to the left due to the certainty effect. Also, this error is magnified as higher probabilities are used in the lottery leading to more risk-averse utility functions (5).

A "paired-gamble" method has been proposed in the literature to account for the certainty effect (4). McCord and deNeufville conducted a study using this paired gamble method, which they called the lottery equivalence method (LE) (5). This method uses elementary lotteries which are binary lotteries with one of the consequences set to the status quo; this reduces the number of parameters to be defined. The subject in the LE method compares the status quo with the assessment probability and also the status quo with the reference probability and provides the point of indifference between these two lotteries. The authors conducted the study with the probability of assesment lottery set at 0.5 and 0.75. The utility of the assessment lottery at the indifference point was calculated as follows:

$$[p(RL) \times U(RL) + [p(SQ) \times U(SQ)] = [p(AL) \times U(AL)]$$

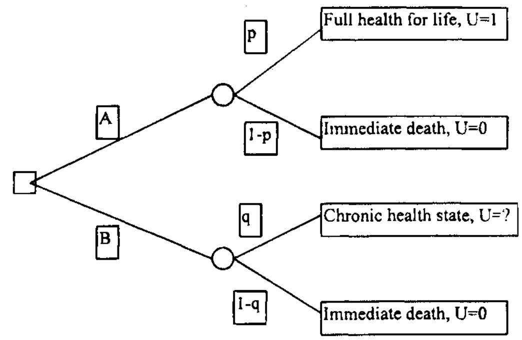
$$+ [p(SQ) \times U(SQ)] \therefore \text{If } p(AL) = q \text{ and } U(SQ) = 0$$

Then, at indifference point

$$U(AL) = [p(RL) \times U(RL)]/q \tag{2}$$

where,

RL = reference lottery,
SQ = status quo,
AL = assessment lottery



At point of indifference,
Expected Utility of A = Expected Utility of B
$\therefore$ [p(FH)×U(FH) + (1-p)(ID)×U(ID)] = [q(CHS)×U(CHS) + (1-q)(ID)×U(ID)]
$\therefore$ p(1) + (1-p) × (0) = q(?) + (1-q) × 0
$\therefore$ U(CHS) = p/q
where, FH=full health, ID=immediate death, CHS=chronic health state

Fig. 2. The lottery equivalence method.

Figure 2 is a representation of the lottery equivalence approach with health states.

They found that use of this method reduced the dependence on the probability and eliminated the certainty effect for positively valued outcomes such as dollars. This finding was confirmed by lower utility values and a utility function which was below and to the right of the utility function for certainty equivalence. They also found that there was no systematic difference between utilities for lotteries with $q=0.5$ and $q=0.75$, further confirming the results.

RATIONALE FOR THE CURRENT STUDY

The literature reports the presence of a bias in utility measurements obtained from the standard gamble method under certainty and probability equivalence, due to the certainty effect (5). Further, the traditional standard gamble method under probability equivalence is used as a "gold standard" in health status utility measurements (10). Based on the literature review and the background provided above, this study had three research objectives:

1. To explore the presence of the certainty effect in health status utility measures obtained using the traditional (PE) standard gamble method.
2. To test the hypothesis regarding the direction of the certainty effect in health state utility measures. The specific hypothesis was that: "the certainty effect in health will occur in a direction opposite to that found by McCord and deNeufville with positively valued outcomes."

The rationale for this hypothesis was that while subjects overweight **positively valued** outcomes occurring under certainty, they would try to avoid the **negatively valued outcome** (i.e., the chronic disease state) which occurred with certainty, and prefer the gamble. Hence the certainty effect would cause individuals to reach indifference between the choices quicker, i.e., at lower reference probabilities. Since the utility of the chronic state depends on the reference probability at indifference point, the utility will tend to be undervalued. Therefore, while the certainty effect led to more risk-averse utility functions with positively valued outcomes, it would lead to more risk-seeking functions (preference for the gamble) with negatively valued outcomes.

3. The third, and final research objective was to explore the possibility of countering the certainty effect using a lottery method of standard gamble.

These ojectives were addressed via a comparison of standard gamble under PE and LE. Assessment probabilities of q=0.5 ($LE_{0.5}$) and q=0.75 ($LE_{0.75}$) were used to make the study comparable to the McCord and deNeufville study. Three chronic disease states were chosen for the study, namely, home dialysis for renal dysfunction, hospital dialysis for the same condition and confinement for a contagious disease. These states were chosen so as to compare the resultant utility values with those reported by Sackett & Torrance in the literature for the same health states by standard gamble (11).

Table 1. Comparison of Mean (±Standard Deviation) Utility Values from OSU Study with Sackett & Torrance (1978)

Health States ⇓	Sackett & Torrance (1978)	OSU (1996)
Contagious disease[a]	n = 192	n = 41
	0.16 ± 0.28	0.28 ± 0.20
Hospital dialysis	n = 189	n = 41
	0.32 ± 0.39	0.44 ± 0.23
Home dialysis	n = 187	n = 41
	0.40 ± 0.42	0.50 ± 0.24

[a] Differences between the two studies were significant at α = 0.05. There were no significant differences in the utility values of the other two health states for the two studies.

MATERIALS AND METHODS

Undergraduate and graduate students of pharmacy in a midwestern university were requested to participate in the study. Fifty-four students volunteered and completed the study. Twenty-one of these were graduate students. Each subject was provided with detailed written descriptions of the three health states which were reproduced from the study by Sackett & Torrance (12). Subjects (Ss) were asked to read the written descriptions carefully and then provide a rank-order preference for these health states by rating scale (13).

Ss were then randomly assigned to receive standard gambles under PE first and LE second (i.e., after 48 hours) or LE first and PE second. A prop called the "Chance Board" designed by Torrance was used to aid Ss in comparing gambles (2,14). The prop was modified to include a lottery for $LE_{0.5}$ and $LE_{0.75}$. The standard gamble was run for each health state by all three methods (PE, $LE_{0.5}$, $LE_{0.75}$) with the order of the gambles from least desirable to most desirable health state as ranked by the subject. Utilities for the three methods were calculated as $U=p$ for PE, $U=p/0.5$ for $LE_{0.5}$ and $U=p/0.75$ for $LE_{0.75}$. Each subject completed all 3 methods, thus the study design was a crossed repeated measures design. A 48-hour "washout" period between the PE and LE methods was used to account for the carryover and learning effects.

RESULTS

Data were analyzed using the SPSSR and SASR systems, versions 6.0. Forty-one of the 54 responses were usable. The rest had utilities of >1.0 in the lottery methods; indicating that these individuals were probably extremely risk-averse; preferring lottery A (in Figure 2) even if it involved a definite percentage of death, in order to avoid the chronic disease state (as was verbalized by some of them). The reason for a moderate number of subjects showing this predilection may be partially attributable to their age group (20–40 years) and thus these disease states may have seemed more "extreme" to them. As stated by Verhoef et al in the context of risk attitudes with life years, ". . . the aspiration level tends to decrease with increasing age, which suggests that it is influenced by realistic expectations about life expectancy." (15)

The percentage of unusable responses could also be attributed to a limitation of the study, i.e., although the subjects were asked to rank-order their preferences for the health states by rating scale, the ranking of the preferences from most to least desired was not independently obtained. It is recommended that this ranking be done prior to the rating scale because there is evidence (16) that there may be a preference reversal between methods if subjects are not asked to explicitly state their rankings, due to the cognitive complexity associated with the relative ordering of several health states.

The study sample comprised 61% females and 39% male undergraduate and graduate pharmacy students. The gender distribution is representative of the pharmacy school enrollment. The statistical power with 41 responses was about 0.9 (at an alpha of 0.05 and a precision of 0.1) (17). All but 8 respondents of the original 54 gave a rank-order of preference as home dialysis > hospital dialysis > contagious disease.

A comparison of the PE utility values with utility values reported in the literature by Sackett & Torrance for the same three health states showed that these were not significantly different at the 95% confidence level, except for the value for contagious disease (Table 1). This result suggests the validity of the standard gamble method under probability equivalence in the current study. It should be noted, however, that the utilities obtained in the present study were uniformly higher than those from the previous study. This could be attributable to the demographics of the subjects, sample size or the change in attitudes towards the health states in the 20 years separating the present study from the earlier one.

Table 2 reports the means (and standard deviations) of the utility values for each health state, obtained via all three methods. The values progressively increased from contagious disease to hospital dialysis to home dialysis for each method, similar

Table 2. Mean (and Standard Deviation) Utility Values for Three Health States Measured by Three Methods

Health state	Method of measuring utility (n = 41)		
	PE[a] 1*	$LE_{0.75}$[b] 2**	$LE_{0.5}$[c] 3**
Contagious disease	0.28 ± 0.20	0.41 ± 0.28	0.51 ± 0.30
Hospital dialysis	0.44 ± 0.23	0.54 ± 0.28	0.62 ± 0.28
Home dialysis	0.50 ± 0.24	0.63 ± 0.25	0.70 ± 0.27

Note: Utility values in columns with similar symbols were not significantly different from each other, those with different symbols were different. All tests were at α = 0.05.

[a] PE = probability equivalence.

[b] $LE_{0.75}$ = lottery equivalence with q = 0.75.

[c] $LE_{0.5}$ = lottery equivalence with q = 0.5.

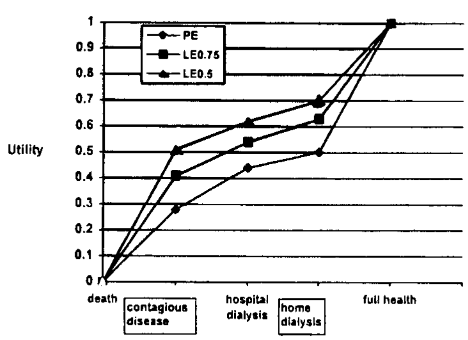
to the trend reported in the literature (11). Data analysis using ANOVA showed that for all three methods, the utility values for the three health states were significantly different from each other using Tukey's multiple comparison method. All analyses were at the 0.05 level.

The results also showed that the utility values for each health state **increased** as the assessment probability **decreased** from $q=1$ (PE) to $q=0.75$ ($LE_{0.75}$) to $q=0.5$ ($LE_{0.5}$), as seen in Table 2 & Figure 3. This finding was exactly the reverse of McCord and deNeufville's findings. This particular finding addressed the first two research objectives, i.e., regarding the presence and direction of the certainty effect in health state utility measures.

A comparison of the utility values obtained by the lottery methods i.e., $LE_{0.5}$ and $LE_{0.75}$, showed that they were not significantly different from each other, but were significantly different from the utility values for PE. This further confirmed that there could be a potential bias due to the certainty effect in PE measurements, which was countered by the lottery methods as indicated by the higher values for both $LE_{0.5}$ and $LE_{0.75}$. The lack of significant difference between the LE methods showed that dependence on the reference probability was being eliminated by the lottery methods. These results addressed the final research objective, which was to explore the possibility of countering the certainty effect using the lottery approach of the standard gamble.

DISCUSSION

The results of this study indicate that the "certainty effect" reported in the utility literature is potentially present in health state utility measurements, albeit in the reverse direction, i.e., it appeared to undervalue utility values for chronic health states which occurred with certainty. This directional effect has been documented by Kahneman and Tversky in the psychology literature as the "reflection effect" (18). Also, as stated by the same authors in their treatise on Prospect theory, this study found that, ". . . the reflection effect implies that risk aversion in the positive domain is accompanied by risk seeking in the negative domain." (18) Further, the lottery equivalence method proposed in the literature appeared to eliminate the bias of the certainty effect as indicated by higher utility values for each health state. The similarity in utility values from the two lottery methods



Note 1: The utility values increased from contagious disease to hospital dialysis to home dialysis, a trend also reported in the literature.

Note 2: The utility values for each health state increased as the assessment probability decreased from 1.0 (PE) to 0.75 ($LE_{0.75}$) to 0.5 ($LE_{0.5}$).

Fig. 3. Plot of utility values versus health states for three measurement methods.

suggested that the "certainty effect" bias was countered by the lottery method. With the lottery equivalence method, Ss showed less risk-seeking and were not so averse to chronic health state lottery as they were to chronic health state certainty. This result could be partially attributed to the possibility of avoidance of the least desirable outcome, i.e., immediate death; a factor explained by "regret theory." (19,20).

Overall, this study showed the presence of differences between the standard gamble methods under PE and LE indicating the possibility of a method effect. These results lead us to suggest that the standard gamble under probability equivalence as it is used in health status assessment, is not invariant with changing parameters and thus the claim of it being the "gold standard" needs further examination.

This study was an initial exploration into this area, and the results are preliminary in nature. The results, however, indicate that further exploration of the phenomenon of "certainty effect" with different parameters such as duration of health state and different probability levels, are warranted. The generalizability of this study would also be enhanced if the study were extended with a larger and more diverse sample of subjects selected from the general healthy, as well as diseased populations.

APPENDIX 1:

The Allais paradox (8)

There are two decisions to be made.

Decision 1 A: Win \$1 million with probability 1
B: Win \$2 million with probability 0.10
Win \$1 million with probability 0.89
Win \$0 with probability 0.01

Choose between A and B at this point.

Decision 2 C: Win \$1 million with probability 0.11
Win \$0 with probability 0.89
D: Win \$2.5 million with probability 0.10
Win \$0 with probability 0.90

Now choose between C and D

It has been noted that about 82% of subjects prefer A over B and 83% prefer D over C. However, this contradicts the normative choice by the maximization of expected utility as is demonstrated below.

Let $U(0) = 0$ and $U(2{,}500{,}000) = 1$ since they are the worst and best outcomes.

$$\text{Then } EU(A) = U(\$1 \text{ million})$$

$$EU(B) = 0.10 + 0.89 * U(\$1 \text{ million})$$

Thus A will be preferred to B if and only if

$$U(\$1 \text{ million}) > 0.10 + 0.89 * U(\$1 \text{ million})$$

or

$$U(\$1 \text{ million}) > 0.91$$

Now, for decision 2,

$$EU(C) = 0.11 * U(\$1 \text{ million})$$

$$EU(D) = 0.10$$

Therefore, D is preferred to C if and only if U($1 million) < 0.91. Since U($1 million) cannot be both greater than and less than 0.91 at the same time, thus choosing A and D is not consistent with expected utility. This inconsistency has been attributed to the fact that outcomes which occur with certainty are overvalued, i.e. the certainty effect.

REFERENCES

1. G. W. Torrance. *J. Chron. Dis.* **40**(6):593–600 (1987).
2. G. W. Torrance. *J. Health Econ.* **5**:1–30 (1986).
3. J. von Neumann and O. Morgernstern. *Theory of games and economic behavior,* 2nd edition, Princeton University Press, New York. 1947.
4. P. H. Farquhar. *Mgmt. Sci.* **30**(11):1283–1300 (Nov. 1984).
5. M. R. McCord and R. deNeufville. *Mgmt. Sci.* **32**(1):56–60 (Jan. 1986).
6. M. Allais. In M. Allais, and O. Hagen, (eds.). *Expected utility hypotheses and the Allais paradox,* D. Reidel Publishing Company, Holland, 1979, pp. 437–482.
7. J. C. Hershey, H. C. Kunreuther, and P. J. H. Schoemaker. *Mgmt. Sci.* **28**(8):936–954 (Aug. 1982).
8. R. T. Clemen. In *Making Hard Decisions: An Introduction to Decision Analysis.* PWS-Kent, Boston, 1991.
9. M. R. McCord and R. deNeufville. *J. Large Scale Sys.* **6**:91–103 (1984).
10. G. W. Torrance. *Socioeconomic Planning Science.* **10**:129–136 (1976).
11. M. F. Drummond, G. L. Stoddart, and G. W. Torrance. In *Methods for the economic evaluation of health care programmes.* Oxford, 1992, pp. 112–148.
12. D. L. Sackett and G. W. Torrance. *J. Chron. Dis.* **31**:697–704 (1978).
13. T. R. Bowe. *Medical Decision Making.* **15**(3):283–285 (Jul–Sep 1995).
14. W. Furlong, D. Feeny, G. W. Torrance, R. Barr, and J. Horsman. Guide to design and development of health-state utility instrumentation. *Working Paper Series.* Ontario, Canada: Center for Health Economics and Policy Analysis, McMaster University, 1990, Paper#90-9.
15. L. C. G. Verhoef, A. F. J. DeHaan, and W. A. J. Van Daal. *Medical Decision Making.* **14**:194–200, 1994.
16. Z. A. Hakim. *Unpublished Ph.D. Dissertation,* The Ohio State University, 1995.
17. R. E. Kirk. *Experimental Design: Procedures for the behavioral sciences.* Second edition. Brooks/Cole, California, 1982.
18. D. Kahneman and A. Tversky. *Econometrica.* **47**:263–291 (1979).
19. G. Loomes and R. Sugden. *Economics Journal.* **92**:805 (1982).
20. G. Loomes and R. Sugden. *J. Econ. Theory.* **41**:270 (1987).