# An Evaluation and Comparison of Utility Measures

for

# Cost Utility Analysis

by

James Andrew Nield
B.A., Bishops University, 1990

A Thesis Submitted in Partial Fulfilment of the
Requirements for the Degree of

MASTER OF ARTS

in the Department of Economics

We accept this thesis as conforming
to the required standard

[redacted]

Dr. J. Schaafsma, Supervisor (Department of Economics)

[redacted]

Dr. J. C. H. Jones, Departmental Member (Department of Economics)

[redacted]

Dr. K.R. Thornton, Outside Member (School of Health Information Science)

[redacted]

Dr. J. Cutt, External Examiner (School of Public Administration)



University of Victoria

# ABSTRACT

Supervisor: Dr. J. Schaafsma

Economic evaluation is gaining prominence in the health care sector as escalating health care costs are forcing policy makers to search for more efficient resource allocation. Cost utility analysis (CUA) is being hailed as a method to compare the marginal costs of a program or treatment with the marginal health improvement attributable to the program or treatment. This health improvement is measured using quality-adjusted-life years (QALY's) gained as the outcome measure in the analysis. Health Status indexes (HSIs) are used to quantify the quality of life component of QALY's. This study compares two methods of generating HSIs, the standard gamble technique (SG) and the time trade-off technique (TTO), to see if they give equivalent results for hypothetical health states in the context of lung cancer treatment.

The equivalence of the SG and TTO is examined using the standard direct comparison as well as a more novel indirect comparison. The direct comparison directly compares the SG and TTO values using regression analysis to test whether they are equivalent. This is the method followed in other studies but it is potentially biased because the dependent variable is limited between 0 and 1. The indirect comparison compares two models relating each of the HSIs to the underlying lung cancer symptoms. The SG and TTO are equivalent if the models relating each HSI to the symptoms are equivalent. This method overcomes the problem of a limited dependent variable because if the coefficients are biased by a limited dependent variable, both models will be biased in the same manner.

Both the direct and indirect comparisons show that the SG and TTO are not equivalent. The indirect comparison also permits some insight into how the two HSIs are different. Specifically, the difference appears to be in only two of the thirty changes in symptom severity levels, when looking at changes between lung cancer health states. Therefore, in most cases, changes in lung cancer health states are valued equivalently. The other important discovery is that the SG varies over demographic groups with different risk attitudes. Males give lower SG values than females, and the females in the student sample give lower values than the females in the sample of clinic workers.

The results of this study have two important policy implications for CUA. First, SG and TTO results are not equivalent. Hence, extreme care should be taken when

comparing the results of CUA studies to ensure that the same HSI is being used. Second, because values for the SG technique vary over demographic groups with different risk attitudes, demographic or cultural differences in the risk attitudes of samples used to generate SG values may bias the results of the study. Therefore, to conduct intra program comparisons using SG values, the demographic composition of the samples must be equivalent.

[redacted]

Dr. J. Schaafsma, Supervisor (Department of Economics)

[redacted]

Dr. J. C. H. Jones, Departmental Member (Department of Economics)

[redacted]

Dr. K.R. Thornton, Outside Member (School of Health Information Science)

[redacted]

Dr. J. Cutt, External Examiner (School of Public Administration)

# ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. J. Schaafsma, for his guidance and effort in seeing me through this endeavour, and the rest of my committee for their insightful comments. I would also like to thank the members of the Lung Cancer Cost Effectiveness Study at the British Columbia Cancer Agency for all their input and the use of the questionnaires obtained from the workers at the Victoria Clinic. Finally, I would like to thank my fiancee, Jenn Vincent, for her love, patience, and commitment throughout this project.

# TABLE OF CONTENTS

ABSTRACT . . . . . ii

ACKNOWLEDGEMENTS . . . . . iv

TABLE OF CONTENTS . . . . . v

LIST OF TABLES . . . . . vii

LIST OF FIGURES . . . . . viii

CHAPTER 1: INTRODUCTION . . . . . 1

CHAPTER 2: LITERATURE REVIEW . . . . . 3
2.1 Background . . . . . 3
2.2 Utility Theory . . . . . 8
2.3 Methods of Utility measurement . . . . . 10
2.3.1 Standard gamble method . . . . . 16
2.3.2 Time trade-off method . . . . . 17
2.3.3 Rating Scale method . . . . . 20
2.4 Equivalence of SG, TTO, and RS HSIs . . . . . 23
2.5 HSIs and sample demographics . . . . . 27
2.6 Outline of Thesis . . . . . 28

CHAPTER 3: DATA . . . . . 30

CHAPTER 4: DIRECT COMPARISON OF SG AND TTO VALUES . . . . . 35

CHAPTER 5: HOW THE SG IS VALUED FROM HEALTH STATES . . . . . 45
5.1 Introduction . . . . . 45
5.2 The model relating SG values to lung cancer health states . . . . . 45
5.3 Testing for age and gender differences . . . . . 48
5.4 Comparison of University and Clinic samples . . . . . 52
5.5 The preferred SG regression equation . . . . . 55
5.6 Interpreting the preferred SG regression results . . . . . 58

CHAPTER 6: HOW THE TTO IS VALUED FROM HEALTH STATES . . . . 61
6.1 Introduction . . . . . 61
6.2 The model relating TTO values to lung cancer health states . . . . . 61
6.3 Testing for age and gender differences . . . . . 62
6.4 Comparison of University and Clinic samples . . . . . 65
6.5 The preferred TTO regression equation . . . . . 68

6.6 Interpreting the preferred TTO regression results . . . . . 70

CHAPTER 7: INDIRECT COMPARISON OF SG AND TTO . . . . . 72

CHAPTER 8: CONCLUSION . . . . . 79
8.1 Overview . . . . . 79
8.2 Summary of the results . . . . . 80
8.3 Conclusions and policy implications . . . . . 83

REFERENCES . . . . . 86

APPENDIX A: QUESTIONNAIRE . . . . . 93

# LIST OF TABLES


Table 1: Direct comparison of the SG and TTO health status indexes 37
Table 2: Direct comparison of the SG and TTO with a truncated sample 41
Table 3: SG regression equations for Clinic and University samples 51
Table 4: SG regression for Combined sample 54
Table 5: SG regressions for preferred model 57
Table 6: TTO regression equations for Clinic and University samples 64
Table 7: TTO regression for combined sample 67
Table 8: TTO regression results for the preferred model 69
Table 9: SG and TTO regressions for indirect comparison 74

# LIST OF FIGURES


Figure 1. Quality and quantity of life 7
Figure 2. Standard gamble method 17
Figure 3. Time trade-off method 18
Figure 4. Wolfson's SG and TTO comparison 25
Figure 5. Wolfson's SG and RS comparison 25
Figure 6. SG and TTO distribution in the University sample 32
Figure 7. SG and TTO distribution in Clinic sample 34
Figure 8. Hypothetical example of how a limited dependent variable can bias OLS coefficients 37
Figure 9. Plot of the SG versus TTO values for the Clinic sample 40
Figure 10. Plot of the SG versus TTO values for the University sample 40

# CHAPTER 1: INTRODUCTION

As competition for health care resources becomes more and more intense, the field of medical decision making is increasingly incorporating economic theory in its search for efficient health care resource allocation. In recent years a major emphasis has been placed on cost effectiveness / cost utility analysis for evaluating health care. In particular, cost utility analysis is becoming prominent in the health care domain because of its ability to compare the marginal costs of a program or treatment with the marginal health improvement attributable to the program or treatment. This health improvement is measured using quality-adjusted-life years (QALY's) gained as the outcome measure in the analysis[1].

To use QALY's there must be a measure of the change in quality of life and quantity of life that is due to the health care program or treatment. Health Status Indexes (HSIs) are used to define the quality of life for different health states. There are three major HSIs that are used to incorporate individuals preferences for different health states. The von-Neumann Morgenstern (1944) standard gamble technique (SG) is the classic method for measuring preferences and is based on the well established foundations of decision theory. The time trade-off technique (TTO) was developed by Torrance (1976) as a simpler to use alternative to the SG but with comparable results. Rating scale techniques (RS) have also gained widespread acceptance because of their simple

[1] Cost utility analysis (CUA) can also been done using healthy year equivalents (HYE's), or other methods of combining the quantity of life and quality of life aspects, as the outcome variable. For the vast majority of cases QALY's are used for CUA.

structure, potential low cost, and historical use in the field of attitude measurement.

Only the SG measure has a solid theoretical foundation to justify its use in economic evaluations. But the SG measure is clinically difficult to use, compared to TTO or RS. However, before the TTO and RS can be used for CUA they must be shown to be acceptable proxies for the theoretically sound SG measure. As shown in chapter 2, below, studies comparing HSIs leave no doubt that rating scales are not equivalent to the SG measure. The results concerning the equivalence of the SG and TTO methods are mixed.

The focus of this paper is on whether the SG and TTO are equivalent HSIs in the context of lung cancer. The SG and TTO values for different lung cancer health states are generated from a questionnaire administered to a sample of University students and a sample of health care professionals at the Victoria Cancer Clinic. A comparison of the SG and TTO indexes can be performed not only by direct comparison but also by comparing models which show how each responds to change in the underlying health state. These models are then compared to see whether the TTO and SG indexes respond in the same way to changes in health states.

In Chapter 2 the HSI literature is reviewed to provide the background and motivation for this thesis. Chapter 3 introduces and explains the data used to compare the SG and TTO HSIs. The direct comparison of the SG and TTO is performed in chapter 4. The models for the SG and TTO are built and examined in chapters 5 and 6, respectively. These models are compared in chapter 7. The conclusions and implications of this study are then discussed in chapter 8.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Background

There is a movement towards economic evaluations in the health care domain as expanding health care needs compete for scarce resources. It is becoming increasingly apparent that the resources for the provision of health care are not unlimited and that the medical community must be cognizant of economic factors when making decisions. As Goodwin (1988) notes "Each time a decision is made that a given treatment should be used, a determination is being made that the available resources are more efficiently used for that treatment then for other treatments." (Goodwin, 1988, 1537). An economic evaluation can help the decision maker to determine the most efficient allocation of scarce resources. This view, that economic evaluations must be undertaken in the health care domain to ensure that resources are being used efficiently, is becoming more prevalent among the medical community. This has led to an intensified interest in using economic methods to evaluate health care programs and treatments.

The two traditional analytic approaches used for examining public programs are cost benefit analysis (CBA) and cost effectiveness analysis (CEA). CBA evaluates projects by calculating the net present value of the costs and benefits in dollar values. The evaluation criterion is based on the 'fundamental rule' "that in any situation involving project choice, the proposal, or group of proposals, to be selected is the one

which produces the greatest net benefit." (Schofield, 1987, 27) In applying this technique to health care there are many problems associated with assigning a monetary value to improved health benefits, improved quality of life, and increased life expectancy. The more prominent methods of assigning dollar values to health care benefits are market valuation, willingness to pay estimates, policy makers views, and professional opinion[2]. Unfortunately, as these various methods give conflicting results, there is no satisfactory method in CBA to assign a monetary value to improved health (Drummond, 1987, 607).

Cost-effectiveness analysis (CEA) avoids the problems encountered in CBA by measuring the impact of a program in non-monetary units only. There is no need to place a monetary value on health outcomes for CEA, because the measure of the health state improvement is simply reported. In CEA the incremental cost of a program is compared to the incremental health effects of the program, where the health effects are measured in natural units related to the objective of the program. If CEA is used to evaluate a program there must be either:

- one, unambiguous, objective of the intervention and therefore a clear dimension along which effectiveness can be assessed.
- or there are many objectives but the alterative interventions are thought to achieve these to the same extent.

Examples of these naturally occurring outcome variables are cases found, cases of disease averted, lives saved, and life years gained. The results are usually expressed in a cost per unit of effect (Drummond et al., 1987). CEA is most useful in comparing programs

[2] For a complete description of these methods see Drummond et al., 1987, 150.

or treatments that have a common outcome variable. CEA comparisons across programs with quite different outcomes are more difficult since these outcome variables are more likely to be measured in different units. There may also be no naturally occurring indicators, such as cases found or life years gained, to reflect the objective of the health state improvement. For example, improving the "quality of life" of patients is the sole objective of palliative treatment for lung cancer. Quality of life is a multidimensional concept with no naturally occurring outcome variables properly reflecting it (Ochs et al., 1988). When intra-program comparison is the goal or when there are no naturally occurring indicators to express the incremental health effects of the program CEA can't be used.

Cost utility analysis (CUA) can be used to perform economic analysis for health care in situations where neither CBA nor CEA are applicable. The incremental health improvement in CUA is a utility measure which focuses on the quality of life, so CUA is not program specific, does not require estimating monetary values, and can be used when quality of life is the important outcome measure. CUA is particularly relevant for lung cancer treatment because the treatment is often aimed solely at improving the patients quality of life (i.e. palliative treatment) and not at a permanent cure. In CUA the incremental cost of a program is compared to the incremental health improvement attributable to the program, where the health improvement is measured in quality adjusted life-years (QALY's) gained. The result is usually expressed as a cost per QALY

gained (Drummond et al., 1987, 112). CUA is much like CEA[3], with the primary difference being that the outcome variable in CUA is a utility measure rather than an objective variable such as life years gained.

The QALY was first introduced by Weinstein and Stason, who defined it as:

> A health-status index is essentially a weighting scheme: each definable health status, ranging from death to coma to varying degrees of disability and discomfort to full health, and accounting for age difference, is assigned a weight from zero to one, and then number of years spent at a given status, Ys, is multiplied by the corresponding weight, λs, to yield a number λs Ys that might be thought of as an equivalent number of years with full health--a number of quality-adjusted life years (QALYs). (Weinstein and Stason, 1977, 718)

QALY's take into account both the quantity of life and the quality of life (i.e. the mortality and morbidity of a treatment), where quality of life is on a 0-1 scale with 1 representing perfect health and 0 representing death. The relationship between quality of life and quantity of life over an individual's lifetime is illustrated in figure 1 for a hypothetical case. The shaded area in figure 1 represents the number of QALY's for the individual's lifetime. To calculate QALY's for use in CUA, the marginal effects of a program or treatment in terms of quality of life and quantity of life must be determined. The quantity of life effect is easily attained by looking at mean survival rate times and survival curves[4]. The issue of quantifying quality of life is not so straightforward. Health

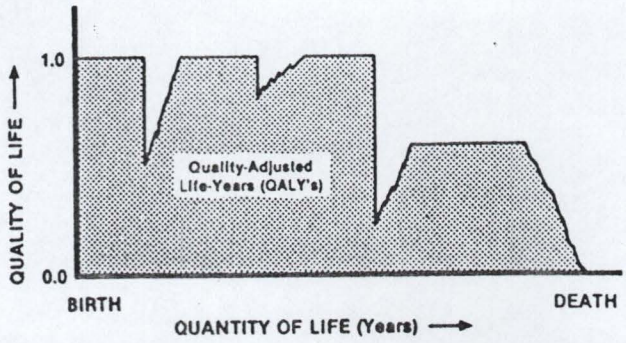
---

[3] It should be noted that some authors choose to use the two names, CUA and CEA, interchangeably.

[4] This aspect is discussed in Whitmore (1976).

status indexes (HSIs) are used to evaluate quality of life and it is critical to CUA that the HSI be accurately measured.

Figure 1. Quality and quantity of life




Source: Torrance, 1987, 593.

The term Utility in CUA originates from the classical method used in developing a HSI, the von Neumann-Morgenstern standard gamble technique (SG), which is based on economic "Utility" theory. This technique and how it can be used to generate HSI values is explained in the next section.

## 2.2 Utility Theory

Economic utility theory is divided into the different but closely related fields of cardinal utility theory and ordinal utility theory. Ordinal utility theory is concerned with individuals ranking preferences while cardinal utility theory actually puts values on those preferences. Cardinal utilities are the type needed for evaluating health states because there is a need to know not only if one health state is preferred to another but by how much it is preferred. The measurement of cardinal utilities (i.e. as in measuring HSIs) is based on the theory of decision making under uncertainty developed by von Neumann and Morgenstern in 1944. The von Neumann-Morgenstern approach is to derive a cardinal utility function for an individual on the basis of that person's preferences between pairs of gambles. Thus, the theory is concerned with decision making in situations in which the outcome is not known with certainty and provides a method by which a person's preferences are revealed through their choice in a series of gambles among different health outcomes (Feeny and Torrance, 1989, s192). The von Neumann-Morgenstern utility function is based directly on six fundamental axioms. In order to represent an individual's preferences over different pairs of gambles as a cardinal utility function a number of assumptions concerning the consistency of the individual's tastes are required.

The six assumptions[5] are presented for an individual's choice between a pair of lottery tickets $L = (p_1A_1, p_2A_2, \ldots, p_rA_r)$ and $L' = (p_1'A_1, p_2'A_2, \ldots, p_r'A_r)$, where a lottery

[5] The formal definition of these assumptions is taken from Luce and Raiffa (1957).

ticket is a chance mechanism which yields the prizes $A_1$, $A_2$, ... , $A_r$ as outcomes with certain known probabilities $p_1$, $p_2$, ... , $p_r$, where each $p_i \geq 0$ and $\sum p_i = 1$. The expression $A_i \geq A_j$ is used to denote $A_i$ is preferred or indifferent to $A_j$. Only minor rewording is needed to fit the utility theory axioms into the context of health states and health state outcomes, specifically the word "prize" must be changed to "health outcomes". The assumption is then that the individual is faced with a finite set of probable health outcomes $A_1$, $A_2$, ... , $A_r$, where a health outcome in this case is a sequence of future health states specified with certainty for the individual.

The six assumptions (axioms) are presented formally but are also clarified in terms of health outcomes in an intuitive manner

Assumption 1 (Ordering of alternatives).

*The "preference or indifference" ordering, $\geq$, holds between any two prizes, and is transitive. Formally, for any $A_i$ and $A_j$, either $A_i \geq A_j$ or $A_j \geq A_i$; and if $A_i \geq A_j$ and $A_j \geq A_k$ then $A_i \geq A_k$.*

Axiom 1 simply states that an individual can preference rank a list of health outcomes from the most preferred to the least preferred (with ties being allowed) and that these rankings are transitive.

Assumption 2 (Reduction of compound lotteries).

Suppose that $L^{(1)}$, $L^{(2)}$, ... , $L^{(s)}$ are any s lotteries which each involve $A_1$, $A_2$, ...

, $A_r$ as prizes. If $q_1, q_2, \ldots, q_s$ are any s non-negative numbers which sum to 1, then $(q_1L^{(1)}, q_2L^{(2)}, \ldots, q_sL^{(s)})$ denotes a compound lottery in the following sense: one and only one of the given s lotteries will be the prize, and the probability that it will be $L^{(i)}$ is $q_i$.

*Any compound lottery is indifferent to a simple lottery with $A_1, A_2, \ldots, A_r$ as prizes, their probability being computed according to the ordinary probability calculus. In particular, if*

$$L^{(i)}=(p_1^{(i)}A_1,p_2^{(i)}A_2, \ldots ,p_r^{(i)}A_r), \quad \text{for } i = 1,2, \ldots ,s,$$

*then*

$$(q_1L^{(1)},q_2L^{(2)}. \ldots ,q_sL^{(s)}) \sim (p_1A_1,p_2A_2, \ldots ,p_rA_r),$$

*where*

$$p_i=q_1p_i^{(1)}+q_2p_i^{(2)}+ \ldots +q_sp_i^{(s)}.$$

Axiom 2 states that any compound lottery (i.e. a lottery in which the payoff is another lottery) can be decomposed into simple lotteries, with the same final outcomes, through the use of probability calculus. For example, consider the compound lottery with a 60% chance of $L^1$ and 40% chance of $L^2$, where $L^1$ is a lottery with a 25% chance of $A_1$ and

a 75% chance of $A_2$, and $L^2$ is a lottery with a 90% chance of $A_1$ and a 10% chance of $A_2$. Assumption 2 states that the compound lottery $(.6L^1, .4L^2)$ is equivalent to the simple lottery $((.6(.25)+.4(.90))A_1, (.6(.75)+.4(.10))A_2)$ which can be simplified to $(.51A_1, .49A_2)$.

Assumption 3 (Continuity).

> *Each prize $A_i$ is equivalent to some lottery ticket involving $A_1$ and $A_r$. That is to say, there exists a number $u_i$ such that $A_i$ is equivalent to $[u_iA_1, 0A_2, \ldots, 0A_{r-1}, (1-u_i)A_r]$. For convenience, we write $A_i \sim [u_iA_1, (1-u_i)A_r] = \tilde{A}_i$, but note well that $A_i$ and $\tilde{A}_i$ are two quite different entities.*

The continuity assumption states that each health outcome is equivalent to some lottery involving only the worst outcome and the best outcome, which for our purposes are death and perfect health, respectively.

Assumption 4 (Substitutability).

> *In any lottery L, $\tilde{A}_i$ is substitutable for $A_i$, that is, $(p_1A_1, \ldots, p_iA_i, \ldots, p_rA_r) \sim (p_1A_1, \ldots, p_i\tilde{A}_i, \ldots, p_rA_r)$.*

The substitutability assumption states that in any lottery a lottery containing the worst outcome (death) and best outcome (perfectly healthy) can be substituted for any outcome.

Assumption 5 (Transitivity).

*Preference and indifference among lottery tickets are transitive relations.*

This assumption strengthens assumption 1, so that the individual can also preference list lottery tickets as well as prizes.

Assumption 6 (Monotonicity).

*A lottery* $[pA_1, (1-p)A_r]$ *is preferred or indifferent to* $[p'A_1, (1-p')A_r]$ *if and only if* $p \geq p'$.

This assumption states that if given two simple lotteries, involving death and perfect health, you would not choose the lottery with the greater probability of death.

In the second edition of their book von Neumann and Morgenstern (1947) prove that if these six axioms hold an individual's preferences can be represented by a cardinal utility function (i.e. a HSI). This is often referred to as the *expected utility hypothesis*:

> *If the preference or indifference relation satisfies assumptions 1 through 6, there are numbers* $u_i$ *associated with the basic prizes* $A_i$ *such that for two lotteries L and L' the magnitudes of the expected values*
>
> $p_1u_1+p_2u_2+ \bullet\bullet\bullet +p_ru_r$ *and* $p_1'u_1+p_2'u_2+ \bullet\bullet\bullet +p_r'u_r$
>
> *reflect the preference between the lotteries. (Keeney and Raiffa, 1976, 29)*

When the six axioms hold an individual's preferences can be represented by a von

Neumann-Morgenstern utility function[6] that is scalar up to a positive linear transformation. This means that this cardinal utility function can be transformed on to any positive linear scale by multiplication and addition to positive constants. For evaluating health state utilities this utility function is put on a 0-1 scale with 0 representing death and 1 representing a normal healthy life.

The focus thus far has been on a cardinal utility function for individuals to value health states. For CUA a societal utility function is desired to evaluate health care programmes. The standard method applied in CUA is to aggregate individual utility functions to get a societal utility function. Because the properties of a von Neumann-Morgenstern utility function are not sufficient for making valid interpersonal comparisons (Luce and Raiffa, 1957, 32) a further assumption is required, specifically[7]:

*Assumption 7*

- *A full, healthy life and death represent all and nothing, respectively, in the health domain and therefore serve as appropriate and well defined reference states for the health utility scale.*
- *Interpersonal utility comparisons are restricted to the health domain.*
- *Within the health domain, all people are equal; that is, a full healthy life should count equally regardless of whose life it is.*

---

[6] In the interest of brevity utility functions mentioned in this paper are of the von Neumann-Morgenstern variety unless otherwise stated.

[7] This assumption is explored in detail in Feeny and Torrance (1989).

This extra assumption allows individual utility values for health states to be aggregated and used as societal utility values for CUA on a 0 - 1 scale. Averaging health state utility values is important because it allows society's utility values for health states to be represented by the average utility value for a representative sample of the population. In generating utility values for health states it is extremely important that samples are representative of the underlying population. This can be done by using large samples and ensuring that there are no systematic biases in samples used. Section 2.5 reviews the findings of previous research in examining how sample make up can affect HSI values.

After showing the theoretical foundations of utility analysis it is only fair to mention some of the objections that have been raised concerning its use. Since a von Neumann-Morgenstern utility function can be shown to exist if the six axioms hold the attacks on this utility function have focused on the validity of the underlying axioms. While at first glance the axioms appear to be reasonable, further scrutiny has revealed some problems. There are many articles pointing out that intransitivities can occur when making paired comparisons.[8] The usual justification that has been presented to counter this problem is that "Although intransitivities occur in practice when making paired comparisons, they are excluded from the theory on the grounds that they are usually measurement errors which can be corrected when pointed out to individuals" (Torrance, 1976, 351). Another problem that has been found is that there can be violations of the

[8] Allais and Slovic (1979) and Lichtenstien (1983), are two of the more prominent articles examining intransitivities when making paired comparisons.

axioms (i.e. inconsistencies in individual's preferences) when dealing with extreme probabilities[9]. This flaw in the axioms is usually interpreted as an inability for individuals to use probabilities close to 0 or 1 rationally, and is taken more as a guide to avoid extreme probabilities rather than a fatal flaw in the axioms.[10] In Torrance's review of utility theory for the evaluation of health care he examines the extent to which these six axioms are valid when applied to the domain of health. He postulates that: "Axiom 3 has been empirically tested in field studies and found to be applicable ... . The other axioms appear to be eminently reasonable when restated in terms of health outcomes" (Torrance, 1976, 353). Torrance, and other practitioners of CUA, are committed to the view that the underlying axioms of utility theory are a firm foundation for a von Neumann-Morgenstern utility function for health states.

While the validity of using cardinal utility values based on a von Neumann-Morgenstern utility function has been challenged, the theory is well established and no alternative theory[11] has managed to replace it as the standard. It will therefore be assumed for the rest of this paper that these axioms do in fact hold for the application of health outcomes.

---

[9] This is referred to as common ratio effects by Allais (1979).

[10] For a further review of the axioms and an analysis of their validity see Luce and Raiffa (1957) or Torrance (1976).

[11] These models fall under the broad category of "generalized expected utility" models. Two of the more popular models are the "prospect theory" of Kahneman and Tversky (1979), and the "regret theory" of Loomes and Sugden (1982).

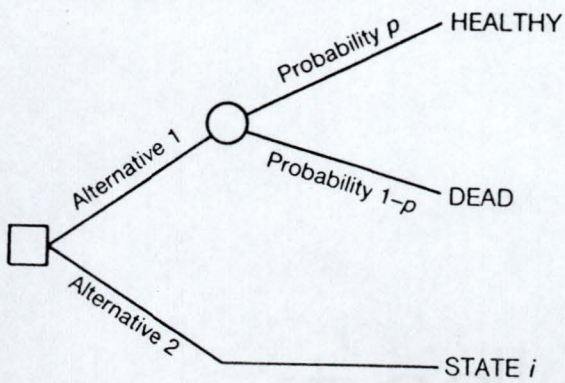
## 2.3 Methods of Utility measurement

### 2.3.1 Standard gamble method

The *classic* method for obtaining preferences based on utility theory is the standard gamble (SG) technique, which is illustrated in figure 2. In the SG technique the subject is offered two alternatives, a certain outcome versus a lottery. The lottery consists of normal health for an additional t years with probability p and immediate death with probability (1-p). The other alternative consists of the certain outcome of a chronic state i for an additional t years[12]. The subject then chooses a probability p so that he or she is indifferent between the two alternatives (i.e. the individual chooses p so that he or she is indifferent between the health state and a lottery with p probability of perfect health and 1-p probability of death). From assumption 7, death was defined as having a health state utility of 0 and perfectly healthy as 1. From the expected utility hypothesis, the magnitude of the individual's preferences can be expressed as $pu_1 + (1-p)u_2$ in a simple lottery. The individual's SG HSI value for the health state is then equal to p because $u_1$ (perfectly healthy) is defined as 1 and $u_2$ (death) is defined as 0.

[12] This method can also be altered to take into account states worse than death and temporary health states.

Figure 2. Standard gamble method




Source: Torrance(1986).

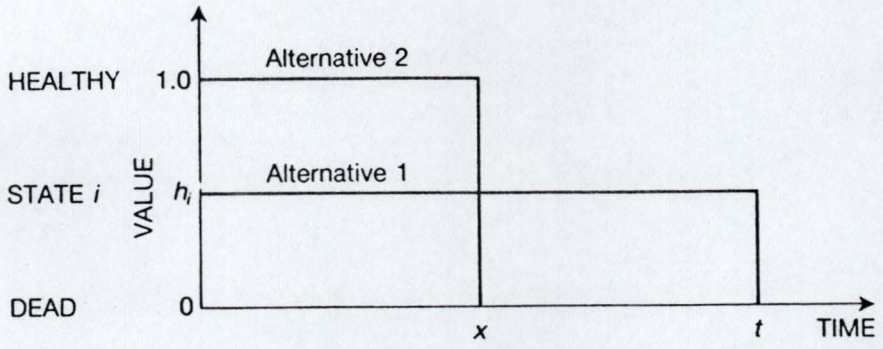
The SG technique is the method outlined by von Neumann and Morgenstern (1947) for representing individual's preferences, after being modified for the health domain. The SG method can be justified for deriving health state utilities, for CUA, on a theoretical basis because the expected utility hypothesis shows it is a linear representation of individuals' preferences if the six underlying axioms hold.

### 2.3.2 Time trade-off method

Torrance et al. (1972) developed the time trade-off (TTO) technique as a simpler alternative to the SG technique. Rather than asking the subject to choose a probability for a gamble between death and perfect health that makes the gamble equivalent to a health state, the subject is asked to trade off remaining time. The procedure elicits preferences or utility values for health states by asking the individual to choose between two

alternatives of certainty. Figure 3 illustrates these two alternatives for a hypothetical health state, $h_i$. Alternative 1 is state i for time t, followed by death. Alternative 2 is good health for time $x < t$, followed by death. The individual is then asked to choose the time x where he or she is indifferent between the two alternatives (Torrance et al., 1972), i.e. $x(1) = qt$, where q is the utility value for health state i. Therefore for the TTO method the utility measure, q, is equal to x/t, the ratio of healthy time to total expected time in health state i.

Figure 3. Time trade-off method




Source: Torrance(1986).

The TTO utility measure differs from the SG in the treatment of risk and time preference. The SG measure incorporates risk or uncertainty into the analysis by measuring a utility value as the probability that makes a lottery between death and perfect health equal to the health state. The uncertainty that is present in the lottery (i.e. a chance of death or perfect health) causes individual's risk attitudes to be incorporated into the SG measure. For example, two individuals, with one having a greater degree of risk

aversion, who judge a particular health state to be equivalent will assign different SG values to the health state. The more risk adverse person will assign a higher SG value for the sole reason of being less willing to accept the chance of death. The TTO measure does not take into account this risk attitude because it deals with certain alternatives. The two measures can be equivalent only if the individuals evaluating the techniques are risk neutral because the uncertainty in the SG technique does not affect the SG values.

The other difference between the SG and TTO methods is the treatment of time preference. The importance of the timing differences between the two methods can be demonstrated by looking at the present values of each technique. The present value of health state i is given by:

$$PV_i = \int_0^t R(t)e^{-rt}\,dt$$

Where R(t) is the health per year on a 0 - 1 sale, t is the remaining years of life, and r is the continuous discount rate for health. The SG equates this health state with a lottery involving a p chance of perfect health for t years and a (1-p) chance of death. The present value of the SG lottery is:

$$PV_{SG} = p\int_0^t (1)e^{-rt}\,dt + (1-p)\int_0^t (0)e^{-rt}\,dt = p\int_0^t e^{-rt}\,dt$$

The TTO method asks individuals to trade off the t years in health state i for x years in perfect health. The PV of the trade off is:

$$PV_{TTO} = \int_0^x (1)e^{-rx}\,dx$$

Setting $PV_{SG} = PV_{TTO}$ and solving for p yields,

$$p = \frac{\int_0^x e^{-rx}\,dx}{\int_0^t e^{-rt}\,dt} \neq \frac{x}{t}$$

The TTO utility measure (x/t) does not equal the SG utility measure (p). If, however, the discount rate for healthy time (r) equals 0, then $PV_{SG} = pt$, $PV_{TTO} = x$, and hence $x/t = p$. For the TTO technique to be equivalent to the SG, the individual's utility from additional healthy years must be a linear function of time, i.e. individuals value a year of healthy time now the same as a year of healthy time 20 years from now (or alternatively that the discount rate for healthy time is zero). It is usually believed that the utility function for healthy time is concave with respect to time (i.e. there is a positive discount rate for healthy time) and the TTO score will underestimate the true utility. This is not important over short time periods where the TTO approximates the SG regardless

of the discount rate for healthy time.

Although the TTO technique is believed to give similar values to the SG[13], it is not derived from the von Neumann-Morgenstern utility model and does not incorporate risk or uncertainty into the analysis. TTO has become very popular for representing health state utilities. It was founded and promoted by Torrance and a group of health economists working out of McMaster University[14], who have spearheaded the movement toward CUA. Although the appropriateness of the TTO method depends on assumptions about the discount rate for healthy time and individual's attitude to risk, it is gaining acceptance and is well worth examining. The TTO method is easier to understand (Torrance, 1976) and is particularly appealing to economists who probably can relate to the concept of trade offs more easily than to probabilities.

### 2.3.3 Rating Scale method

The third type of HSI is the rating scale (RS) which is sometimes referred to as category scaling. A RS consists of a line or Likert scale with clearly defined endpoints. One endpoint is the most preferred health state and the other is the least preferred health state. Individuals then rate health states on the scale as a direct measure of their preference for the health state. To use a RS for CUA these endpoints are 0 (death) and

---

[13] This belief is based on the previous comparisons of the two HSIs and is discussed in section 2.5.

[14] This group includes Sackett, Drummond, Thomas, Stoddard, and Feeny.

1 (perfectly healthy) and other states are rated linearly as to where they are put on the rating scale. Rating scales have their roots in psychometrics[15] and are in widespread use in clinical analysis. They are preferable to other HSIs in terms of cost, simplicity, and cohesiveness with established medical practices. A rating scale is intuitively simple and does not require any knowledge of utility theory on the part of individuals evaluating health states and the person organizing the study. A rating scale does not involve the difficult question of trading off a health state with a lottery involving perfect health and death, or trading off time for a particular health state. It is therefore easier to complete because it does not tax the individuals assessing the health states as heavily as the SG or TTO.

Theoretically, the problem with the rating scale is that it does not explicitly take the individuals preferences into account for each health state. The health states are given a value between 0 and 1 but the concept of gambling or trading off that health state for another is not dealt with in this method. The rating scale ranks the health states but does not give an actual utility value for each health state. There is no theoretical reason to believe that the RS method gives cardinal utility values. It is therefore not surprising that evaluations have shown RS values to differ from those of the SG and TTO[16].

---

[15] Torgerson (1958) and Baird et al. (1978) are reference texts for an overview of psychometrics.

[16] Studies comparing the RS to the TTO and SG are presented in the next section.

## 2.4 Equivalence of SG, TTO, and RS HSIs

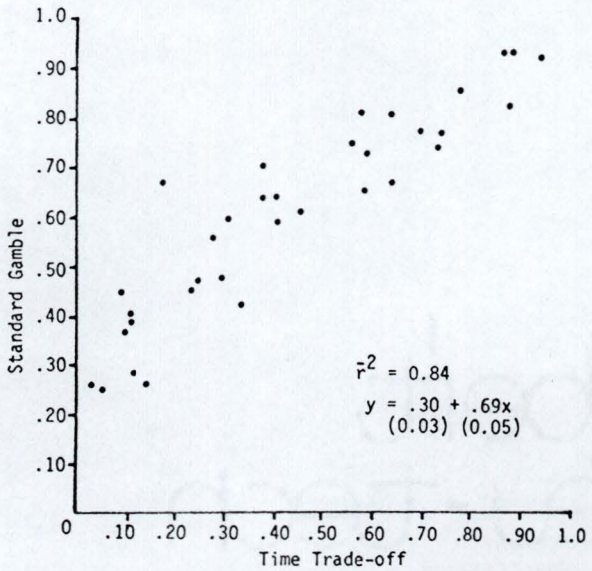
A fundamental issue in the HSI literature is whether the TTO and RS yield reliable HSIs. Only the SG has a sound theoretical base, so it is critical to examine how the TTO and RS HSIs compare to the SG measure. Due to the popularity and attributes of the TTO and RS it is imperative to know if these methods are justified for CUA. The literature to date has generated mixed results concerning the compatibility of these HSIs.

Torrance (1976) was the first to compare the three measurement techniques using a representative cross section of the general public. Regression analysis was used to compare the SG on TTO HSIs, by regressing SG on the TTO. The equivalence of these indexes is examined by testing whether the constant in the estimated equation is equal to zero and the slope coefficient is equal to one. The estimated regression results are SG = .11 + .83 TTO, with $R^2$ = .95. In generating these results Torrance uses average HSI values for each health state rather than using the HSI values for each individual in his regressions. This causes the test results to become somewhat suspect due to the extremely small sample size (i.e. N=6). There is a 0.65 correlation coefficient between the TTO and SG values, which Torrance reports as satisfactory. It is debatable whether two variables with a correlation coefficient of 0.65 can be considered equivalent. Based on his view that the SG and TTO methods are equivalent Torrance goes on to evaluate the RS by comparing it to the TTO method. He found that the fit was very poor when the TTO was regressed on the RS and that the correlation between the two is only 0.36. Torrance's conclusion regarding the comparability of the three methods is that SG and

TTO methods are equivalent and the SG and RS are not. Although Torrance and others take it for granted that the SG and TTO methods are equivalent, it is debatable whether this is justified, based on his results. For a small set of health states studied, the mean values for the SG were very close to the TTO values, but this does not prove conclusively that the two methods are equivalent.

Wolfson et al. (1982) also compared the results of TTO and RS to that of the SG technique using a group of stroke patients as the sample. Regression analysis was again used to examine the comparability of the three HSIs. It was found that neither TTO nor RS had the expected coefficients when regressed on the SG (i.e. a constant of 0 and a slope of 1). This can be seen by examining their scatter plots and their estimated regression equations, reproduced as figure 4 and 5 below.

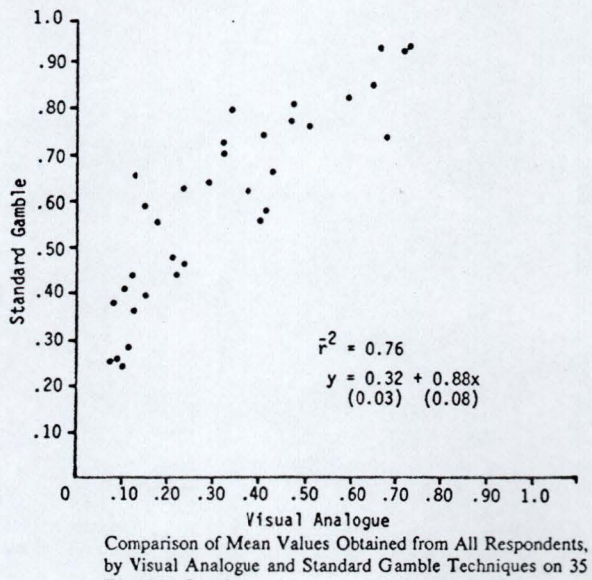
Figure 4. Wolfson's SG and TTO comparison




Comparison of Mean Values Obtained from All Respondents, by Time Trade-Off and Standard Gamble Techniques on 35 Disability Levels

Source: Wolfson et al., 1982, 208.

Figure 5. Wolfson's SG and RS comparison




Comparison of Mean Values Obtained from All Respondents, by Visual Analogue and Standard Gamble Techniques on 35 Disability Levels

Source: Wolfson et al., 1982, 207.

Although the $\overline{r}^2$ values for both equations are relatively high Wolfson concludes that the SG, TTO, and RS are not equivalent. One reason is because in the estimated regression equations of the TTO and RS regressed on the SG the constant term is not equal to zero and the slope is not equal to 1, indicating that the RS and TTO are not equivalent to the SG. It is also noted that the SG and TTO technique produced 17 statistically significant differences in 35 health states rated, while the SG and RS produced 33 statistically significant differences in 35 states rated. Wolfson concludes that neither the TTO or RS are equivalent to the SG and that TTO and RS are the two most closely related HSIs.

A study by Read et al. (1984) used analysis of variance to test the null hypothesis that SG, TTO, and RS produce equivalent preference scales. This experiment was conducted using a sample of physicians attending a seminar on preference assessment methodology. They reported that the SG, TTO, and RS produced different scale values for outcomes of a clinical problem.

There are also studies that compare two of the HSIs. Quin (1981) found that there were significant differences between RS and SG responses in a medical decision making task using college students as subjects. O'Connor et al. (1987) and Llewellyn-Thomas et al. (1982) both found SG values to be significantly higher than RS values. A group based at the University of California and headed by Bush (1973) conducted studies comparing rating scales to the TTO method and found them to be comparable. While Churchill et al. (1987) found only a low correlation (0.22) between TTO and RS.

In the review of past results on the comparability of the three methods it seems

there is a general consensus that the RS and SG methods are not equivalent while there are mixed views on the comparability of the SG and the TTO method. Torrance (1986) is the only one to report that the SG and TTO are equivalent and these conclusions may be unfounded. This lack of a clear solution to whether the TTO is equivalent to the SG is the underlying motivation for the comparison of these two HSIs.

## 2.5 HSIs and sample demographics

It is the prevalent view in CUA that the demographics of the sample do not affect HSI values (Drummond et al., 1987, 118). This view, which is very convenient to CUA research, is based on the work of Churchill et al. (1987), Wolfson et al. (1982), and Kaplan and Bush (1978). Their research shows HSI values to be stable over samples with different age, sex, income, education, and medical experience.

There are a few studies that show HSIs to be affected by the demographics of the sample involved. For example, Sackett and Torrance (1978) found age to be a significant influence in the way health states are evaluated, while socio-economic status and sex didn't affect the way HSIs are evaluated. It should be noted that none of these studies performs a thorough analysis on the SG HSI. Due to the unique attribute of the SG measure incorporating risk and uncertainty into the analysis it could possibly be different than the other HSIs that do not incorporate risk. Demographic or cultural groups that have different risk attitudes would also value the SG differently for health states. The TTO technique incorporates a time preference for healthy time into its HSI values. So

if demographic or cultural groups have a different time preference for healthy time, they would also have different TTO values.

## 2.6 Outline of Thesis

This thesis addresses the issue of whether or not the TTO can be used as a proxy for the SG HSI in CUA. Although the TTO has no theoretical basis it is commonly used for CUA on the grounds that it provides equivalent HSI values to the SG and is much simpler to use. As shown in the review of previous results, there is no substantive empirical evidence showing that the SG and the TTO are in fact equivalent, leaving a need for further research in this area.

The equivalence of the SG and TTO is examined using a direct comparison method and an indirect comparison method. The direct comparison is the traditional method for comparing HSIs and involves comparing the SG and TTO values directly. The indirect comparison is performed to overcome some econometric problems inherent in the direct comparison. The indirect comparison involves estimating the parameters of a model that relate HSI values to various lung cancer health states, where the HSI values are first generated using the SG technique and then the TTO technique. If the two HSIs are equivalent, the parameter estimates of the two models for the HSIs should also be equivalent.

In this study three demographic characteristics are accounted for: age, gender, and medical knowledge. These demographic groups may have different risk attitudes, and/or

preferences for healthy time and hence may assign different SG and TTO values to a given health state. As was indicated in the previous section, research in this area is sparse, so it is informative to test whether these demographic differences affect the valuation of the health states. Although many other demographic or cultural groups may have different time preferences for healthy time or risk attitudes, this analysis is limited to groups represented in the samples.

# CHAPTER 3: DATA

This chapter describes how the SG and TTO data were obtained and gives an overview of the data sets. The SG and TTO HSI values were generated with a questionnaire developed by Dr. J. Schaafsma and C. Auld for use in a fourth year economics honours thesis. A copy of the questionnaire is attached as appendix A. This questionnaire is designed to elicit utility values for different health states associated with lung cancer.

Each questionnaire describes six health states (a page for each), and a cover page which asks the individual's age and sex and explains how to evaluate the health states using the SG and TTO approaches. Each health state is described by different severity levels for seven symptoms related to lung cancer. These seven symptoms and severity levels were taken verbatim from the EORTC questionnaire (Aaronson et al., 1991), which evaluates the important aspects of the quality of life for lung cancer. The seven symptoms are the most representative symptoms out of a list of thirteen of the relevant symptoms for lung cancer patients. The severity levels for each of the seven symptoms were generated randomly, for each health state, and each questionnaire was randomly assigned six health states. The severity level for each symptom can take on one of four possibilities: "not at all", "a little", "quite a bit", and "very much". The symptoms used in the questionnaire are:

1. hair loss

2. vomiting
3. depression
4. shortness of breath while resting
5. trouble swallowing
6. chest pain
7. coughing blood

Each health state, consisting of different severity levels for these symptoms, was assumed to have a time span of twenty years followed by death.

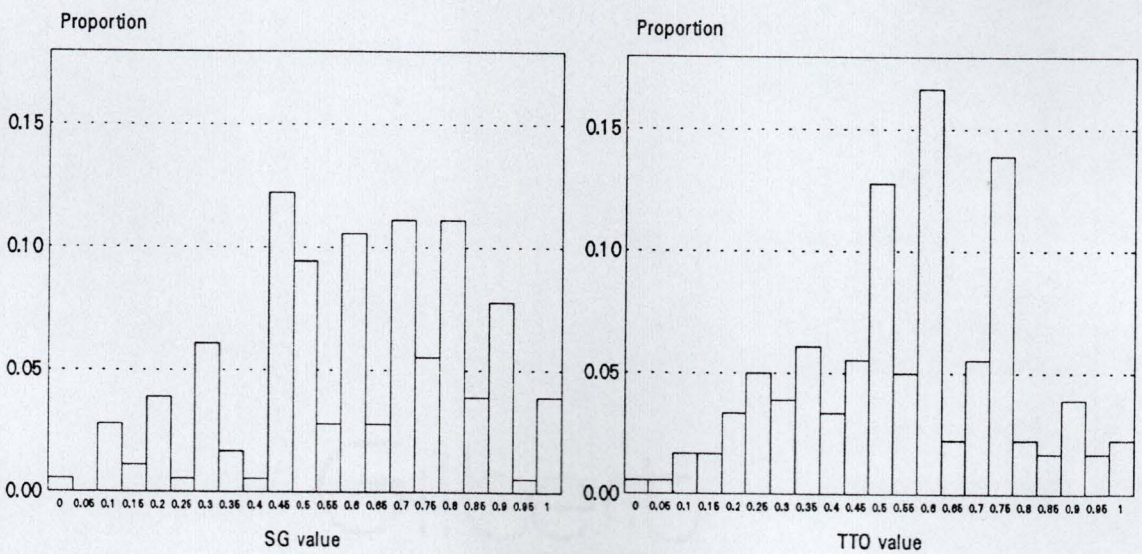
The subjects were asked to imagine that they are in the ill health state as described by the severity of the symptoms, and that they have a life expectancy of twenty years. The SG instrument is presented as a drug that, if successful, gives full health for the twenty years life expectancy. If the drug is unsuccessful immediate death results. The subject is asked to choose the minimum probability of success (p) that the drug must have for him or her to be willing to take it. This p is the SG value for the health state.

The TTO instrument is presented as a different drug that provides a number of years (x), between 0 and 20, in full health with certainty. The subject is asked the minimum number of years in full health that the drug has to provide for he or she to be willing to take it. The TTO value for the health state is simply x/20.

The questionnaire was given to two different samples of subjects. The two samples were chosen to determine if medical knowledge of lung cancer health states affects the way in which the SG and TTO are valued. The first sample is a first year principles level economics class at the University of Victoria. This sample is generally

not familiar with health states associated with lung cancer so they are dealing with hypothetical cases. For the student sample the questionnaire was given to a class of approximately 45 students and 35 questionnaires, with six health states each, were returned. Of the 35 questionnaires returned 5 were unusable. Two were incomplete, and three were either confused or did not take due care in filling out the forms. One student had a constant utility across health states that varied greatly and the other two had alternating 1 and 0 utility values for no apparent reason. By pure coincidence there were 15 males and 15 females in the sample. The age in the sample ranged from 17 to 36 but was clustered around 20. All of the symptoms are evenly distributed between the severity levels since the symptoms are randomly generated. The SG has a mean value of .59 and the TTO has a mean value of .55. The distribution of the SG and TTO values are shown in figure 6.

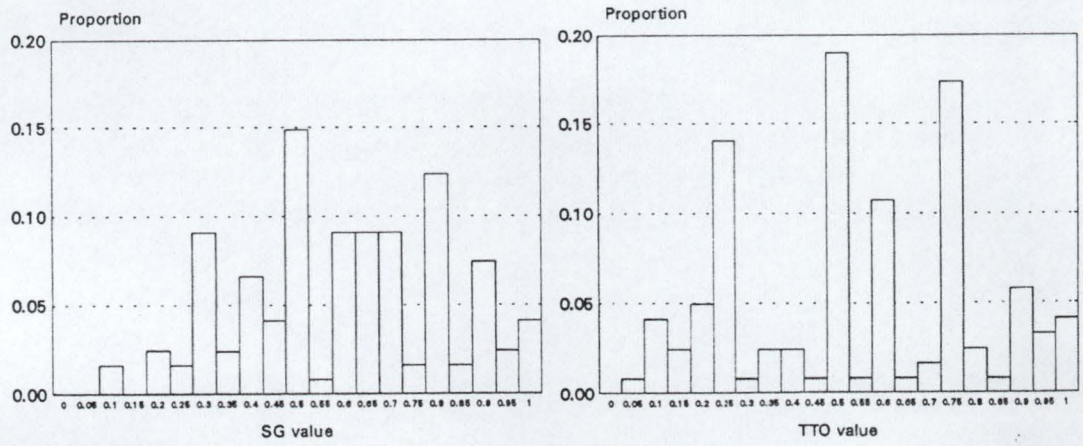
Figure 6. SG and TTO distribution in the University sample



The shape of the of the SG and TTO distributions are quite similar in general, but the actual utility value numbers with the highest proportion of values are different for the two HSIs. For example, for the TTO the .75 value has a large proportion while the .7 value does not, but for the SG this is reversed.

The other sample consists of a group of health care professionals at the British Columbia Cancer Agency, Victoria Clinic. This would be the preferred sample because their medical knowledge of the health states related to lung cancer enables the Clinic workers to better understand the true meaning of the symptom severity levels. In the Clinic sample the same questionnaire with the addition of an extra cover page to explain the questionnaires relationship to the Clinic, shown in appendix A, was given on a volunteer basis to workers at the Clinic. There were 25 questionnaires returned with 20 of the questionnaires being usable. Four questionnaires were incomplete and one person misunderstood the questionnaire, having TTO values moving in the opposite direction of the SG. In this sample there were 6 males and 19 females. The ages in this sample were more or less uniformly distributed from 27 to 57. All of the symptoms are evenly distributed between the severity levels since the symptoms are randomly generated. The SG has a mean value of .60 and the TTO has a mean value of .54. The distribution of the SG and TTO values are shown in figure 7.

Figure 7. SG and TTO distribution in Clinic sample




For the Clinic sample the shape of the distribution is similar for the two HSIs, but the SG is more evenly distributed over the scale. That is to say for the TTO most of the values are either .25, .5, .6, or .75. Of the 20 questionnaires returned by the Clinic staff 2 are from medical staff, 2 are from nursing staff, 7 are from technical staff, 2 are from health records, 4 are from the other category, and 3 had no profession indicated. Because of the relatively small size of the Clinic sample the information on the respondent's profession could not be used.

The surveys were given to both samples on a voluntary basis. Since there is no a priori reason to believe that those who didn't complete the questionnaire are different from those who did, there is no need to be concerned about sample bias.

# CHAPTER 4: DIRECT COMPARISON OF SG AND TTO VALUES

The justification for using the TTO technique rather than the theoretically rigorous SG technique is that it is a simpler technique with equivalent results. The hypothesis that TTO and SG are equivalent can be tested directly using multiple regression analysis and the two data sets. A direct comparison can be performed by regressing the TTO values on the SG values and testing the functional form and goodness of fit. In the direct comparison test, SG is regressed on TTO using the linear specification shown in equation (DC-1).

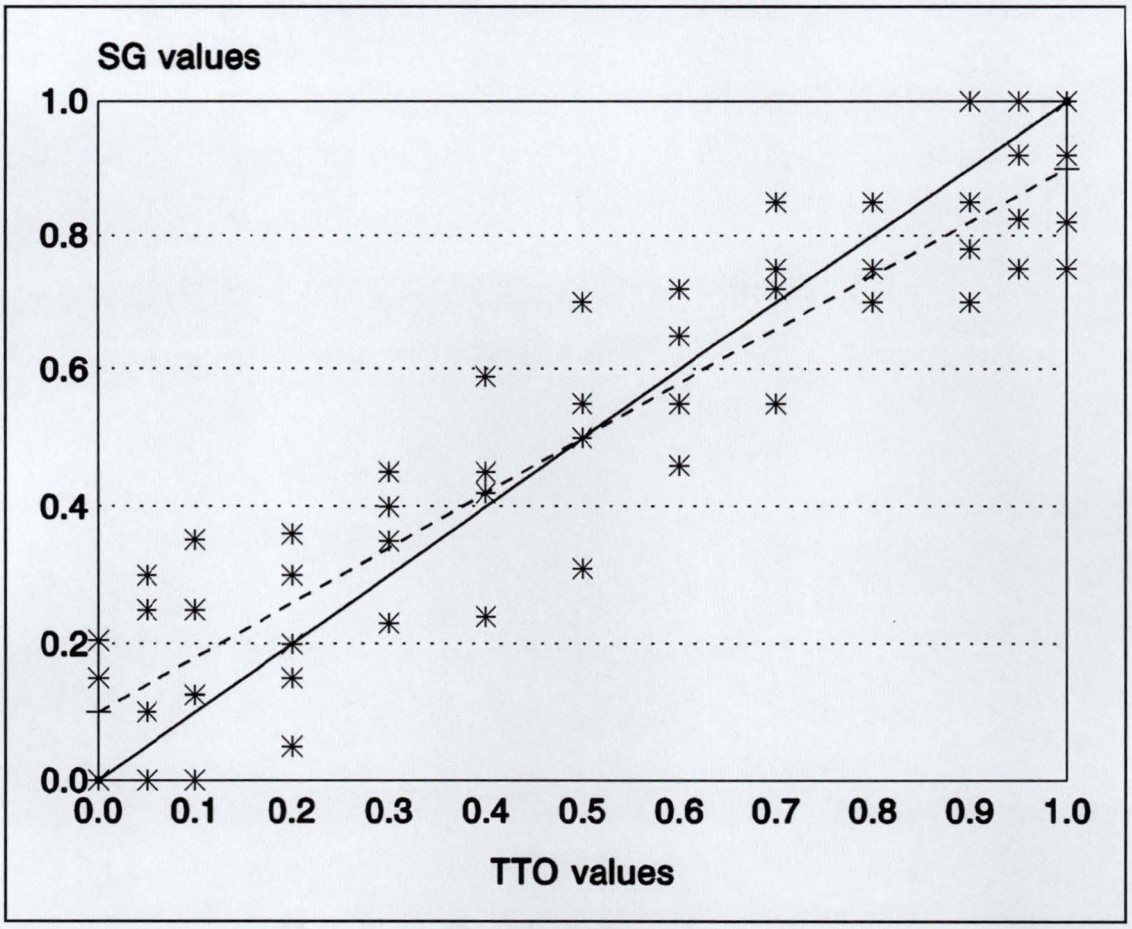
$$SG = \alpha + ß\ TTO + \varepsilon_1 \qquad \text{(DC-1)}$$

Where $\varepsilon_1$ is the error term in the equation. If TTO is equivalent to the SG, $\alpha$ should be equal to 0 and ß should be equal to 1. Therefore, a joint hypothesis test is conducted to test this hypothesis.

When performing this analysis there are econometric issues that must be taken into account. Due to the dependent variable's limited nature (i.e. bounded between 0 and 1), there may be a problem with the error terms being skewed when the dependent variable is near 0 or 1, resulting in biasing the coefficients to indicate that $\alpha > 0$ and $ß < 1$. Figure 8 represents how the limited nature of the dependent variable, SG, can

bias the estimated coefficient. For this hypothetical example the true equation is SG = TTO, so the constant is equal to zero and the slope is equal to 1 (i.e. the SG and TTO are equal at every point). The data points in figure 8 are hypothetical SG / TTO values for health states. Two of the assumptions for OLS are that the mean value for the error term is zero and that there is homoscedasticity or equal variance for the error terms. These two assumptions are necessary for OLS to provide Best Linear Unbiased Estimators (BLUE). As can be seen in figure 8, when the SG values are near zero or one, both assumptions are violated. Because the SG can not go lower than 0 or greater than 1, the error terms near the endpoints are skewed (i.e. not equally dispersed around the true regression line). OLS minimizes the sum of the squared residuals, and can be shown (i.e. the Gauss Markov theorem) to give BLUE estimates if the underlying assumptions hold. Unfortunately since the error terms are skewed near the endpoints OLS gives biased results. Minimizing the sum of squares in this case will rotate the estimated regression clockwise. Since this problem only occurs in samples that test health states with SG values near zero or one, studies that do not include any extreme health states can avoid this problem. Some of the mixture of results in the literature could be explained if some of the samples contained extreme health states and other samples did not. This problem may be identified by testing for heteroscedasticity and normality of the error terms. A possible correction for this is to omit sample points where the TTO variable is near 1 or 0.

Figure 8. Hypothetical example of how a limited dependent variable can bias OLS coefficients




Where:

—— is the true regression line
------ is the biased regression equation
* are the hypothetical data points

Although there may be some statistical problems associated with the direct comparison it gives an indication of how the two HSIs compare. This is the method that

is used in previous studies (i.e. Torrance (1976) and Wolfson et al. (1982)) for comparing the two HSIs and gives a useful reference point for comparison with earlier studies.

The regression results obtained by estimating equation DC-1 using the Clinic and University samples separately, and the results of the joint test of $\alpha=0$ and $ß=1$ are presented in table 1.

Table 1: Direct comparison of the SG and TTO health status indexes

| | UNIVERSITY SAMPLE | | CLINIC SAMPLE | |
|---|---|---|---|---|
| **VARIABLE** | **ESTIMATED COEFFICIENT** | **T RATIO** | **ESTIMATED COEFFICIENT** | **T RATIO** |
| TTO | 0.8116 | 15.48 | 0.6826 | 8.84 |
| Constant | 0.1435 | 4.60 | 0.2352 | 8.79 |
| $\overline{R}^2$ | 0.5714 | | 0.6449 | |
| N | 180 | | 121 | |
| Wald test for $\alpha=0$ and $ß=1$ (with 2 DF) | 24.79* | | 75.93* | |
| B-P-G ($\chi^2$) test for Heteroscedasticity | 1.808 | 1 df | 1.788 | 1 df |
| GF ($\chi^2$) test for normality of residuals | 25.32 | 11 df | 6.16 | 8 df |

* There is significant evidence at the 1 % level to reject the null hypothesis.

The Wald tests in table 1 show that the null hypothesis, that $\alpha = 0$ and $ß = 1$ (i.e SG = TTO), is rejected, at the 1% level of significance, for both of the samples. Due

to the limited nature of the dependent variable (i.e. it is bounded between 0 and 1), the possibility that the results are biased must be examined. A potential bias could be due to the non normality of the error term (i.e. the error term may be skewed near the SG endpoints because the SG values may not be above 1 or below 0). A Wald test is used because it does not require the normality of the error term for large samples. Even though the Wald test is unbiased when the error terms are not normally distributed, the test can still be prejudiced if the coefficients that are being tested are biased. Therefore, the limited dependent variable may cause the null hypothesis of SG / TTO equivalence to be falsely rejected.

It is the skewed error terms at the end of a limited dependent variable that can bias the results. These skewed error terms may be detected by testing for heteroscedasticity and non normality of the error terms. The Breush-Pagan/Godfrey statistics in table 1 shows there is no evidence of heteroscedasticity for the Clinic and University sample, at the 10% level of significance. The Goodness of Fit statistic indicates only the University sample tests positive for non-normal residuals at the 1% level of significance. There is a mixture of results with neither sample testing positive for heteroscedasticity but the University sample does indicate non normal error terms. The limited dependent variable could be influencing the estimated regression results in the university sample.

Even if the estimated regression results do not test positively for heteroscedasticity the regression results can still be biased. This is because there is heteroscedasticity on a priori grounds (i.e. because of the limited dependent variable) even though there could

be no heteroscedasticity around the fitted line. So it is not known whether the results of the tests are biased or not. Figures 9 and 10 show the plot of SG versus TTO for the Clinic and University samples, respectively.

Figure 9. Plot of the SG versus TTO values for the Clinic sample

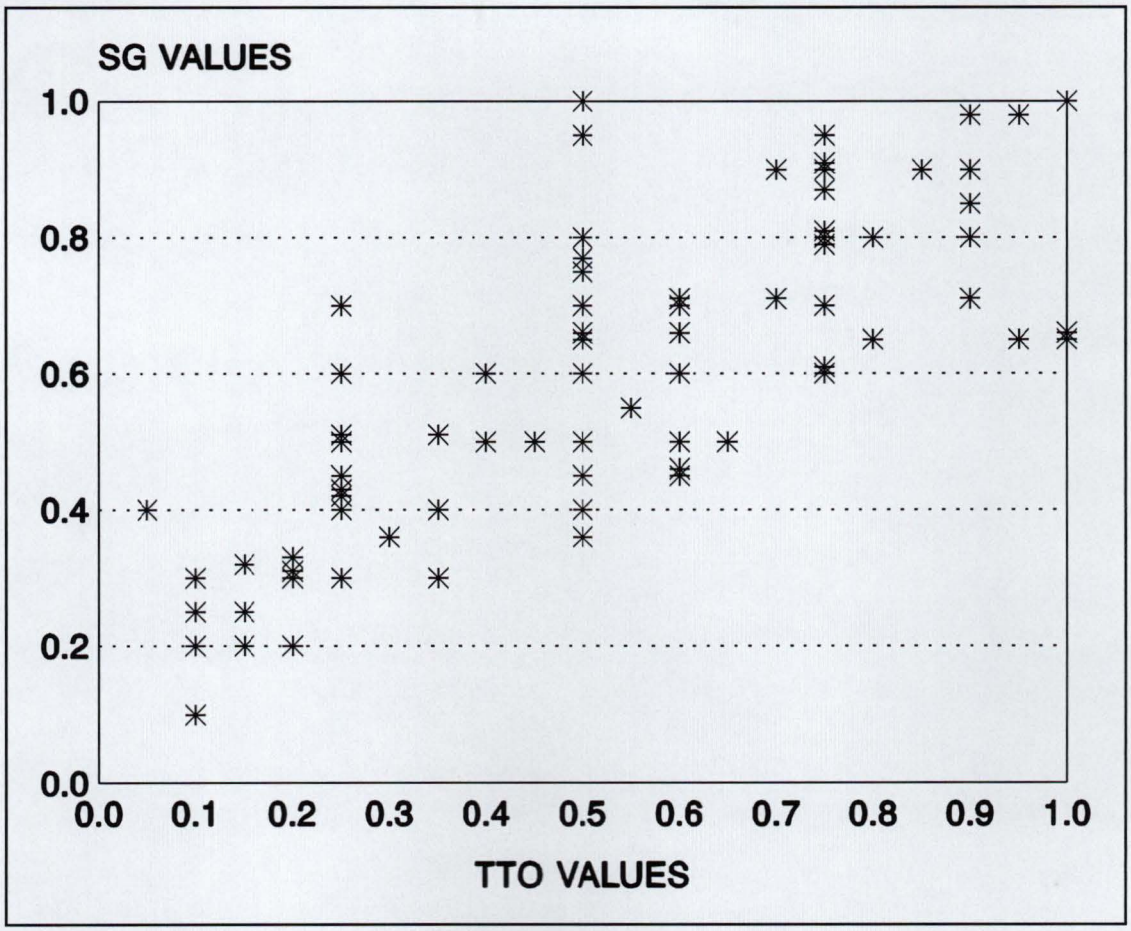
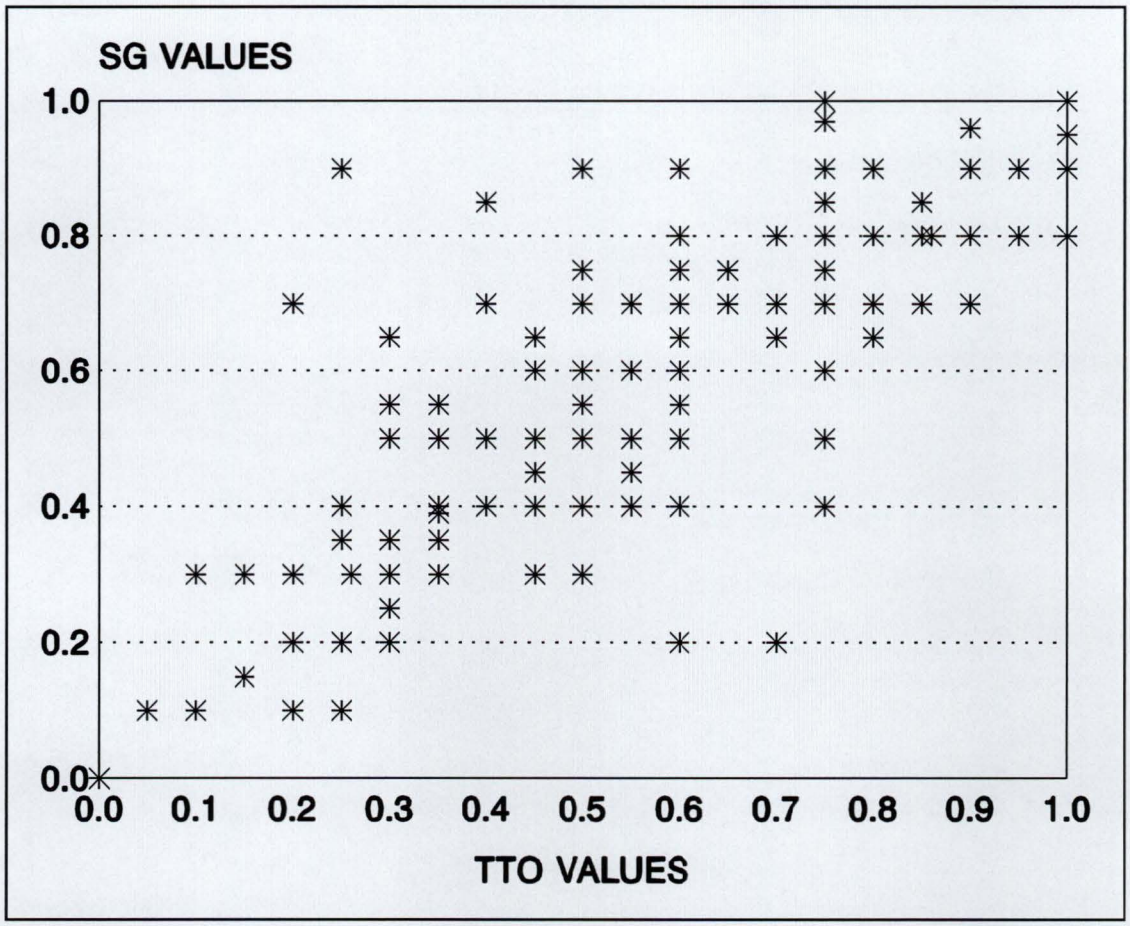


Figure 10. Plot of the SG versus TTO values for the University sample




The scatterplots in figures 9 and 10 appear very similar. Both have a majority of their values in the middle range (i.e. between .2 and .8). As is expected there is a much larger dispersion around the middle than at the endpoints. Figure 8, for the University sample, has a greater disparity between the middle and the endpoints, which may explain why this sample tested positively for non-normality of the residual and the other sample did not. It is hard to judge whether the limited nature of the SG variable is responsible

for biasing the coefficients shown in table 1. This is because there are not many data points where the SG values are very close to the endpoints.

To ensure that the results have not been biased by the econometric problems that were encountered in estimating equation (DC-1), the root of these problems must be eliminated. Since it is only at the limits for the SG values that the error terms could be skewed, if the values for the TTO which cause the SG to be at the limit are omitted then the problems may be resolved. Equation DC-1 is re-estimated for the two samples with the TTO values restricted to values greater than 0.15 and less than 0.85. Dropping the TTO values from 0 to 0.15 and 0.85 to 1.0 eliminates most of the extreme SG values without reducing the sample size too much. As can be seen in figures 8 and 9, if the top and bottom 15% of the TTO are cut off then there are very few SG points at the extremities and the limited nature of the SG variable should not affect the results. The results for these regressions and the tests performed on them are summarized in table 2.

Table 2: Direct comparison of the SG and TTO with a truncated sample

| | UNIVERSITY SAMPLE | | CLINIC SAMPLE | |
|---|---|---|---|---|
| **VARIABLE** | **ESTIMATED COEFFICIENT** | **T RATIO** | **ESTIMATED COEFFICIENT** | **T RATIO** |
| TTO | 0.8270 | 11.23 | 0.7826 | 12.82 |
| Constant | 0.1422 | 3.48 | 0.2015 | 6.13 |
| $\overline{R}^2$ | 0.4450 | | 0.6275 | |
| N | 157 | | 98 | |
| Wald test for $\alpha=0$ and $ß=1$ (with 2 DF) | 21.52* | | 70.56* | |
| B-P-G ($\chi^2$) test for Heteroscedasticity | 0.642 | 1 df | 0.012 | 1 df |
| GF ($\chi^2$) test for normality of residuals | 18.94 | 11 df | 7.54 | 2 df |

* There is significant evidence at the 1 % level to reject the null hypothesis.

In both truncated samples $\alpha$ moves closer to 0 and ß moves closer to 1, but the results of the estimation do not change much. The Wald test still rejects the null hypothesis that $\alpha=0$ and $ß=1$ at the 1% level of significance for both samples. Both samples test negatively for heteroscedasticity and positively for non normality of the error terms. So, even though the extreme values for the TTO are removed there is still a problem of non normal residuals. But, cutting the endpoints off does not seem to affect the results of the test. This indicates that the rejection of the hypothesis that the SG and TTO are equivalent is not due to a limited dependent variable problem.

The above results can be used to make some initial judgements on the equivalence of the two HSIs. First, the rejection of the null hypothesis is very significant even when

the endpoints are excluded, so it is hard to believe that the limited nature of the dependent variable could be the sole reason for such a strong rejection. These results indicate that the SG and TTO are not equivalent, but the limited dependent variable problem leaves a cloud of suspicion over the results.

Because of the uncertainty of the results obtained by using the traditional direct comparison of the SG and TTO measures, other methods of comparison must be attempted to overcome these problems. One such method is the indirect comparison. The indirect comparison compares the two models relating the SG and TTO utility measures to the seven lung cancer symptoms. The SG and TTO are equivalent if the models relating each HSI to the symptoms are equivalent. This method overcomes the problem of limited dependent variables because if the coefficients are biased by a limited dependent variable, both models will be biased in the same manner. So limited dependent variables will not affect the test of whether or not the two models are equivalent. The indirect comparison thus overcomes the uncertainty of the direct comparison.

# CHAPTER 5: HOW THE SG IS VALUED FROM HEALTH STATES

## 5.1 Introduction

In this chapter a model relating SG utility values to lung cancer health states is developed. The SG values are examined as a function of the seven lung cancer symptoms. The demographic factors of age, gender, and medical knowledge are also taken into account.

## 5.2 The model relating SG values to lung cancer health states

The model developed and estimated in this chapter relates the SG measure to the seven symptoms that make up the lung cancer health states. Each symptom has four severity levels (i.e. "not at all", "a little", "quite a bit", and "very much"). Because of their categorical nature the symptoms can not be expressed as linear variables. For example, there is no a priori reason to believe that a move from "not at all" to "a little" is the same as a move from "quite a bit" to "very much". The appropriate method to account for the categorical nature of these independent variables is to include a dummy variable[17] for each of the categories for each symptom.

---

[17] A dummy variable is a variable that is created for each category. It is equal to one for that category and zero otherwise.

This is shown in model SG-1.

$$\begin{aligned} SG = {} & \alpha_0 + \alpha_1 \text{ Hairloss}_1 + \alpha_2 \text{ Hairloss}_2 + \alpha_3 \text{ Hairloss}_3 + \alpha_4 \text{ Vomiting}_1 \\ & + \alpha_5 \text{ Vomiting}_2 + \alpha_6 \text{ Vomiting}_3 + \alpha_7 \text{ Depressed}_1 + \alpha_8 \text{ Depressed}_2 \\ & + \alpha_9 \text{ Depressed}_3 + \alpha_{10} \text{ Breathing}_1 + \alpha_{11} \text{ Breathing}_2 + \alpha_{12} \text{ Breathing}_3 \\ & + \alpha_{13} \text{ Swallowing}_1 + \alpha_{14} \text{ Swallowing}_2 + \alpha_{15} \text{ Swallowing}_3 + \alpha_{16} \text{ Chest Pain}_1 \\ & + \alpha_{17} \text{ Chest Pain}_2 + \alpha_{18} \text{ Chest Pain}_3 + \alpha_{19} \text{ Coughing Blood}_1 \\ & + \alpha_{20} \text{ Coughing Blood}_2 + \alpha_{21} \text{ Coughing Blood}_3 + \varepsilon_2 \end{aligned} \qquad \text{(SG-1)}$$

The three dummy variables for each symptom are dichotomous variables for the severity level of the symptom. For $\text{symptom}_j$, j represents the severity level of the symptom, j=1 for "not at all", j=2 for "a little", and j=3 for "quite a bit". The fourth category is left out, so "very much" for each of the seven symptoms is the base category in the model. Therefore, the coefficient for each of the symptom variables, represents the change in the SG value due to a move in that symptom from "very much" to the severity level j, with the other symptoms held constant. The coefficients for the symptoms are expected to be positive and to increase in magnitude as j goes from 3 to 1, for each $\text{Symptom}_j$. This is because the worst case is left out, so every other severity level, all other things being equal, is expected to have a larger SG value.

It is reasonable to expect people to differ in the way that they convert health states

into SG values[18]. Model SG-1 can not be used because the random error term, $\varepsilon_2$, has a systematic component that is unaccounted for in this model. The previous model is modified to account for the variation between individuals, by including interpersonal dummy variables to account for the variation in the constant that occurs between the k subjects[19]. This specification is shown in model SG-2[20].

$$SG = \alpha_1 \text{ Hairloss}_1 + \alpha_2 \text{ Hairloss}_2 + \alpha_3 \text{ Hairloss}_3 + \alpha_4 \text{ Vomiting}_1 + \ldots$$
$$+ \alpha_{21} \text{ Coughing Blood}_3 + \beta_1 \text{ IP}_1 + \ldots + \beta_k \text{ IP}_k + \varepsilon_3 \qquad \text{(SG-2)}$$

The interpersonal dummy variables $IP_j$ are dummy variables that = 1 if it is subject j and 0 otherwise. There is no constant included in this model because the interpersonal variables are in fact the constant for each individual (i.e. the coefficient for $IP_i$ is the constant for the $i^{th}$ person in the study). An aggregate model, such as SG-1, is obtained by taking the average of the coefficients for the IP variables to get an average constant.

In examining how SG values are generated from health states, four different tasks are performed: first, each of the samples is tested for age and gender differences; second, the University and Clinic samples are compared; third, the two samples are combined where it is appropriate; and fourth, the results for the combined samples are discussed

---

[18] CUA acknowledges these differences and accounts for them by using mean values for health states.

[19] The variation in the coefficient between individuals can not be accounted for because of the limited size of the data.

[20]This is the same methodology that is used in models that incorporate both time series and cross-sectional components (Green, 1990, 483).

and examined. These tasks are performed using Model SG-2.

## 5.3 Testing for age and gender differences

The SG measure implicitly incorporates risk into its valuation of health states, implying that individuals with different risk attitudes assign different SG values. Because age and gender groups may differ in their risk aversion, they may also vary in assigning SG values. There are no previous studies showing that SG values are not affected by age and gender so these factors must be tested for.

SG-2 allows for differences in the constant term (i.e. each individual is given a different constant). This model is estimated for the Clinic and University samples (i.e. table 3) and at the 1% level of significance the average constants for males and females are different for both samples[21]. For both samples the average male constant is lower than the average female constant. This indicates that males accept a higher probability of death in their gambles and hence are less risk adverse. Due to the difference in the average constants for males and females, these constants must be presented separately when SG-2 is estimated. This male / female difference is an important result for CUA because it indicates that the sample used to elicit SG values for health states must contain an appropriate male / female mix.

The hypothesis that the coefficients for the severity level of symptoms (i.e. $\alpha_1$ through $\alpha_{21}$) are different for males and females is also tested. This is done by first

[21]All the joint tests conducted are Wald tests unless otherwise specified.

creating a dummy variable for gender, MALE, that is equal to one if the person is male and zero otherwise. A multiplicative dummy variable is then created for each of the 21 variables by multiplying the male variable by each of these variables. This specification is shown in SG-3.

$$\begin{aligned} SG = {} & \alpha_1 \text{ Hairloss}_1 + \tau_1 \text{ (MALE)*Hairloss}_1 + \alpha_2 \text{ Hairloss}_2 \\ & + \tau_2 \text{ (MALE)*Hairloss}_2 + \ \ldots \ + \alpha_{21} \text{ Coughing Blood}_3 \\ & + \tau_{21} \text{ (MALE)*Coughing Blood}_3 + \beta_1 \text{ IP}_1 + \ldots + \beta_k \text{ IP}_k + \varepsilon_4 \end{aligned} \qquad \text{(SG-3)}$$

$\tau_1$ through $\tau_{21}$ are the coefficients for the multiplicative dummy variables for gender and symptom severity level. Each of the $\tau_i$ coefficients represents the difference between males and females for the $\alpha_i$ coefficient in equation SG-2. A statistically significant $\tau_i$ implies that the coefficient for the symptom severity level, $\alpha_i$, is different for males and females.

At the 5% level of significance the 21 coefficients for the symptom severity levels are not different for males and females in both the Clinic and University samples (i.e. $\tau_1$ through $\tau_{21}$ are jointly not different from zero). Individually none of the coefficients for change in severity level are significantly different from zero at the 5% level (i.e. all of the $\tau$'s are insignificant) for both the Clinic and University samples. So, males tend to give lower SG values than females, but still interpret changes in the severity levels of symptoms in the same manner.

Age is tested in the same manner as gender[22] and is not a significant determinant in the valuation of SG health states. Due to the lack of variation in age in the University sample, the effect of age on the standard gamble value for a given health state can only be tested on the Clinic sample. Even then, there are only 20 individuals in the sample and the results are not conclusive.

The regression results for equation (SG-2) are shown in table 3, for the Clinic and University samples. Because the average male and female constants were found to be different they are presented separately in the table.

[22] Age is turned into a dichotomous variable by dividing the sample into age groups and assigning dummy variables for the different groups.

Table 3: SG regression equations for Clinic and University samples

| | CLINIC SAMPLE | | UNIVERSITY SAMPLE | |
|---|---|---|---|---|
| **VARIABLE*** | **ESTIMATED COEFFICIENT** | **T RATIO** | **ESTIMATED COEFFICIENT** | **T RATIO** |
| $Hairloss_1$ | 0.0519 | 1.25 | 0.0483 | 1.39 |
| $Hairloss_2$ | 0.0924 | 2.01 | 0.0118 | 0.03 |
| $Hairloss_3$ | 0.0630 | 1.57 | -0.0026 | -0.08 |
| $Vomiting_1$ | 0.0711 | 2.10 | 0.1481 | 2.48 |
| $Vomiting_2$ | 0.0744 | 1.80 | 0.0797 | 2.47 |
| $Vomiting_3$ | 0.0516 | 1.24 | -0.0140 | -0.41 |
| $Depressed_1$ | 0.0062 | 0.16 | 0.1488 | 4.34 |
| $Depressed_2$ | 0.0433 | 1.17 | 0.1862 | 5.13 |
| $Depressed_3$ | 0.0179 | 0.45 | 0.0998 | 2.75 |
| $Breathing_1$ | 0.0977 | 2.16 | 0.0438 | 1.34 |
| $Breathing_2$ | 0.0728 | 1.72 | 0.0444 | 1.27 |
| $Breathing_3$ | 0.0991 | 2.24 | -0.0064 | -0.20 |
| $Swallowing_1$ | 0.0354 | 0.93 | 0.0569 | 1.73 |
| $Swallowing_2$ | 0.0318 | 0.75 | 0.0219 | 0.64 |
| $Swallowing_3$ | 0.0226 | 0.59 | -0.0262 | -0.78 |
| Chest $Pain_1$ | 0.1452 | 3.34 | 0.1542 | 4.54 |
| Chest $Pain_2$ | 0.1001 | 2.42 | 0.0679 | 1.84 |
| Chest $Pain_3$ | 0.0518 | 1.16 | 0.0842 | 2.57 |
| Coughing $Blood_1$ | 0.0694 | 1.74 | 0.1051 | 2.97 |
| Coughing $Blood_2$ | 0.0772 | 1.80 | 0.1588 | 4.27 |
| Coughing $Blood_3$ | 0.0253 | 0.59 | 0.0390 | 1.25 |
| Average Constant** | 0.2786 | | 0.2322 | |
| Male Constant** | 0.1701 | | 0.1998 | |
| Female Constant** | 0.3148 | | 0.2646 | |
| $\overline{R}^2$ | 0.6419 | | 0.6429 | |
| Deg of Freedom | 80 | | 129 | |

* For $Symptom_j$, the severity level is "not at all" for j=1, "a little" for j=2, "quite a bit" for j=3, while the base case is "very much".

** The constants are the average constants for the individuals in each group.

## 5.4 Comparison of University and Clinic samples

In developing HSI values for CUA it is recommended that a sample with knowledge of the health states involved be used (Drummond et al., 1987). Health care professionals in the treatment of lung cancer have a good understanding of the symptoms and are the preferred sample to be used for valuing lung cancer health states. The HSI values for the two different samples used in this paper can be used to see if there is a difference between the way the preferred sample (i.e. the health care professionals at the Victoria cancer Clinic) and a sample of convenience (i.e. University students) value the health states.

Table 3 shows the estimated regression results for SG-2, for the Clinic and University samples. These results can be used to see how the coefficients for the estimated models of SG-2 are different for the two samples. But statistical tests must be performed to determine whether these differences are statistically significant. To test whether the Clinic and University results are equivalent, the basic model must be re-estimated with the two samples combined into one sample. The average constants for males are not significantly different for the University and Clinic samples, at the 5% level. The average female constants are however different in the two samples at the 5% level of significance.

To test whether the coefficients for the symptom severity levels are different for the two samples a dummy variable, STUDENT, is created that is equal to one if the individual is a student and zero otherwise. The student variable is then multiplied by each of the 21 variables for the symptom severity levels. This specification is shown as SG-4.

$$SG = \alpha_1 \text{ Hairloss}_1 + \tau_1 \text{ (STUDENT)*Hairloss}_1 + \alpha_2 \text{ Hairloss}_2$$
$$+ \tau_2 \text{ (STUDENT)*Hairloss}_2 + \ \ldots \ + \alpha_{21} \text{ Coughing Blood}_3$$
$$+ \tau_{21} \text{ (STUDENT)*Coughing Blood}_3 + ß_1 \text{ IP}_1 + \ldots + ß_k \text{ IP}_k + \varepsilon_4 \quad \text{(SG-4)}$$

In a joint test, the estimated coefficients for the 21 multiplicative dummy variables (i.e. $\tau_1$ through $\tau_{21}$) are not significantly different form zero at the 5% level of significance. Individually, the coefficients for the Depression variables are the only symptom coefficients that are different in the two samples at the 5% level of significance. So changes in severity levels of all the symptoms but Depression are interpreted in the same manner for the two samples when assigning SG values. As can be seen in table 3 the University students have much higher coefficients for depression than the Clinic workers, indicating that the students interpret depression as a much more important factor in their valuation of health states.

With a few specification adjustments a single equation can be estimated with a combined student and clinic sample. The adjustments consist of allowing the coefficients on the Depression variables to be different for student and clinic workers and presenting the average female constants for each sample. Depression(student) is equal to Depression*STUDENT, where STUDENT is as defined for SG-4. Depression(Clinic) is equal to Depression*CLINIC, where CLINIC is a dummy variable equal to 1 if the individual is in the clinic sample and zero otherwise. This model is estimated for the combined sample in table 4.

Table 4: SG regression for Combined sample

| | COMBINED SAMPLE | |
|---|---|---|
| **VARIABLE*** | **ESTIMATED COEFFICIENT** | **T RATIO** |
| $Hairloss_1$ | 0.0477 | 1.81 |
| $Hairloss_2$ | 0.0448 | 1.63 |
| $Hairloss_3$ | 0.0197 | 0.77 |
| $Vomiting_1$ | 0.1142 | 4.86 |
| $Vomiting_2$ | 0.0766 | 3.04 |
| $Vomiting_3$ | 0.0135 | 0.52 |
| $Depressed(student)_1$ | 0.1454 | 4.40 |
| $Depressed(student)_2$ | 0.1780 | 5.17 |
| $Depressed(student)_3$ | 0.0983 | 2.83 |
| $Depressed(clinic)_1$ | 0.0396 | 1.02 |
| $Depressed(clinic)_2$ | 0.0609 | 1.66 |
| $Depressed(clinic)_3$ | 0.0503 | 1.28 |
| $Breathing_1$ | 0.0570 | 2.24 |
| $Breathing_2$ | 0.0500 | 1.90 |
| $Breathing_3$ | 0.0302 | 1.19 |
| $Swallowing_1$ | 0.0369 | 1.51 |
| $Swallowing_2$ | 0.0310 | 1.22 |
| $Swallowing_3$ | -0.0105 | -0.42 |
| Chest $Pain_1$ | 0.1581 | 6.06 |
| Chest $Pain_2$ | 0.0901 | 3.30 |
| Chest $Pain_3$ | 0.0687 | 2.68 |
| Coughing $Blood_1$ | 0.0864 | 3.34 |
| Coughing $Blood_2$ | 0.1195 | 4.37 |
| Coughing $Blood_3$ | 0.0368 | 1.49 |
| Male constant** | 0.1893 | |
| Female Constant(student)** | 0.2543 | |
| Female Constant(clinic)** | 0.3302 | |
| $\overline{R}^2$ | 0.6347 | |
| Deg of Freedom | 227 | |

* For $Symptom_j$, the severity level is "not at all" for j=1, "a little" for j=2, "quite a bit" for j=3, while the base case is "very much".

** The constants are the average constants for the individuals in each group.

## 5.5 The preferred SG regression equation

While the estimated regression equation in table 4 generally conforms to "a priori" expectations there are some insignificant variables and unexpected coefficient patterns for the severity levels of some symptoms. Specifically, for the Hairloss symptom only the coefficient for the "not at all" severity level is statistically significant at the 5% level. Also for this variable there is no significant evidence to state that the coefficients for "not at all" and "a little" are different at even the 90% level. When interpreting the Hairloss symptom, individuals are distinguishing between two categories, one includes the "not at all" and "a little" severity levels, and the other includes the "quite a bit" and "very much" severity levels. The "quite a bit " variable can be combined with the "very much" (i.e. by omitting it from the equation) because people are not distinguishing between it and "very much". The "not at all" and "a little" categories for Hairloss can be combined because there is no distinction between these severity levels for hair loss.

For the Vomiting symptom the coefficient for the "quite a bit" severity level is not statistically significant at the 5% level. The "very much" and "quite a bit" severity levels can therefore be combined by omitting the "quite a bit" variable, so the base case combines the two highest severity levels.

There is no statistically significant difference between the "not at all" and "a little" coefficients for depression at the 10% level for the students. These two variables can be combined into one variable that represents both severity levels jointly. For the

Clinic workers there is no statistically significant difference, at the 10% level, between the coefficients for the "not at all" and "a little" severity levels for depression. When the three severity levels are represented by one variable it is statistically significant at the 10% level.

For each of the Breathing, Swallowing, and Coughing Blood symptoms the coefficients for "not at all" and "a little" are not statistically different at the 10% level, so the "a little" and "not at all" severity levels can be represented by one variable. The "quite a bit" coefficient is not statistically significant at the 10% level and can be omitted.

The regression results for the model incorporating these changes are presented in table 5.

Table 5: SG regressions for preferred model

| | COMBINED SAMPLE | |
|---|---|---|
| **VARIABLE*** | **ESTIMATED COEFFICIENT** | **T RATIO** |
| $Hairloss_{12}$ | 0.0406 | 2.23 |
| $Vomiting_1$ | 0.1074 | 5.10 |
| $Vomiting_2$ | 0.0719 | 3.17 |
| $Depressed(student)_{12}$ | 0.1558 | 5.38 |
| $Depressed(student)_3$ | 0.0972 | 2.87 |
| $Depressed(clinic)_{123}$ | 0.0544 | 1.99 |
| $Breathing_{12}$ | 0.0467 | 2.71 |
| $Swallowing_{12}$ | 0.0400 | 2.22 |
| Chest $Pain_1$ | 0.1577 | 6.18 |
| Chest $Pain_2$ | 0.0831 | 3.23 |
| Chest $Pain_3$ | 0.0650 | 2.62 |
| Coughing $Blood_{12}$ | 0.0820 | 4.41 |
| Male Constant** | 0.2331 | |
| Female Constant(student)** | 0.2973 | |
| Female Constant(clinic)** | 0.3694 | |
| $\overline{R}^2$ | 0.6411 | |
| Deg of Freedom | 239 | |

* For $Symptom_j$, the severity level is "not at all" for j=1, "a little" for j=2, "quite a bit" for j=3, while the base case is "very much".

** The constants are the average constants for the individuals in each group.

## 5.6 Interpreting the preferred SG regression results

In this section the estimated regression equation in table 5 is examined to determine how individuals are assigning SG values to health states. The estimated coefficients in the equation represent the change in SG value that is due to a change in the severity level of that symptom from the base severity level, "very much", to the severity level of the variable in question, holding the severity levels of all the other symptoms constant. These coefficients are examined to determine how the severity levels are interpreted for each of the symptoms, and to judge the relative importance of the symptoms in assigning SG values to lung cancer health states.

The Chest Pain symptom is the only symptom for which each severity level has its own unique effect on the SG value. As the Chest Pain severity level declines (i.e. moves from "very much" to "not at all") a higher SG value is assigned for each of the severity levels. For Hairloss, Breathing, Swallowing, and Coughing Blood people are not distinguishing between the worst two severity levels (i.e. "very much" and "quite a bit") and are also not distinguishing between the best two severity levels (i.e. "not at all" and "a little"). These symptoms are therefore interpreted in a dichotomous manner (i.e. as either bad or good). For the Vomiting symptom people interpret the "quite a bit" and "very much" severity levels as equivalent, showing that they view quite a bit of vomiting to be equally bad as very much vomiting. However, people do distinguish between the "not at all" and "a little" severity levels for vomiting, so there are three categories of severity levels that are interpreted; "not at all", "a little", and a joint category for "quite

a bit" and "very much". For Depression the University students interpret a little and no depression to be equivalent (i.e. a little depression is not an important factor). The Clinic workers don't distinguish between the "not at all", "a little", and "quite a bit" severity levels for depression. So Clinic workers don't even view "quite a bit" of depression as an important symptom.

The relative magnitude of the coefficients indicates how important each symptom is in assigning a SG value to health states. Since all the coefficients are dummy variables, representing a change in SG value due to a change in the symptom from the "very much" to the severity level in question, a comparison of the coefficients is possible. The coefficients to be examined are presented in table 5. Depression(student)$_{12}$ and Chest Pain$_1$ are the two largest coefficients (i.e. approximately .15), while Vomiting$_{12}$ and Coughing Blood$_{12}$ also have relatively large coefficients (i.e. .1074 and .0820 respectively), and all the other symptoms have coefficients under .065. It is the chest pain and depression (student sample only) symptoms that are the most important symptoms when people attach a SG value to health states. This is extremely useful information when considering how to evaluate treatment options for lung cancer in a cost utility framework. Because the chest pain and depression (student sample only) symptoms give the highest changes in SG utility values for decreases in their severity level, treatment should be aimed at alleviating these symptoms. For example, since a decrease in the severity level of chest pain from "very much" to "not at all" gives an accompanying increase in utility value 3 times greater than an identical decrease in the severity level for the breathing symptom,greater emphasis should be placed on treating

the chest pain symptom. Because the coefficients in table 5 represent the importance people assign to the severity levels of the symptoms when determining the quality of life, they can be used to determine which symptoms should receive treatment priority.

The average male constant is lower than the average female constant, and the average student female constant is lower than the average Clinic female constant. This indicates that males assign lower SG values to the lung cancer health states than females, and that student females assign lower values than the Clinic females. A possible explanation of this male / female / clinic / student difference is that males are less risk adverse than females and that the student females are less risk adverse than the Clinic females. Because the SG technique incorporates this risk attitude, the three groups give different SG values. The implication of this result for CUA is that SG values can vary, depending on what sample is used to generate these values.

# CHAPTER 6: HOW THE TTO IS VALUED FROM HEALTH STATES

## 6.1 Introduction

In this chapter the analysis presented in chapter 5 is repeated with the TTO utility measure, rather than the SG measure, as the dependent variable. The TTO results are presented in the same manner as the SG results, but in a somewhat abbreviated form to avoid repetition. Particular emphasis is placed on results that are different than those presented for the SG.

## 6.2 The model relating TTO values to lung cancer health states

The model used to relate the TTO utility measure to the seven symptoms that make up lung cancer health states is the same as the model presented for the SG and is shown as equation TTO-1 below,

$$\begin{aligned} TTO = {} & \alpha_1 \text{Hairloss}_1 + \alpha_2 \text{Hairloss}_2 + \alpha_3 \text{Hairloss}_3 + \alpha_4 \text{Vomiting}_1 \\ & + \alpha_5 \text{Vomiting}_2 + \alpha_6 \text{Vomiting}_3 + \alpha_7 \text{Depressed}_1 + \alpha_8 \text{Depressed}_2 \\ & + \alpha_9 \text{Depressed}_3 + \alpha_{10} \text{Breathing}_1 + \alpha_{11} \text{Breathing}_2 + \alpha_{12} \text{Breathing}_3 \\ & + \alpha_{13} \text{Swallowing}_1 + \alpha_{14} \text{Swallowing}_2 + \alpha_{15} \text{Swallowing}_3 + \alpha_{16} \text{Chest Pain}_1 \\ & + \alpha_{17} \text{Chest Pain}_2 + \alpha_{18} \text{Chest Pain}_3 + \alpha_{19} \text{Coughing Blood}_1 \\ & + \alpha_{20} \text{Coughing Blood}_2 + \alpha_{21} \text{Coughing Blood}_3 \\ & + ß_1 IP_1 + \ldots + ß_k IP_k + \varepsilon_4 \end{aligned} \qquad \text{(TTO-1)}$$

where TTO is the utility value for the TTO technique and all the other variables are defined as before. Since the worst case is the base case the coefficients for the symptoms are expected to be positive and to increase in magnitude as j goes from 3 to 1.

In examining how TTO values are generated from health states, four different tasks are performed; first, each of the samples is tested for age and gender differences; second, the University and Clinic samples are compared; third, the two samples are combined where it is appropriate; and fourth, the results of the combined sample are discussed and examined. These tasks are performed using Model TTO-1 to interpret how TTO values are generated from the lung cancer health states using the same methodology that is employed in the previous chapter.

## 6.3 Testing for age and gender differences

The TTO technique implicitly incorporates individual's time preferences for healthy time into its valuation of health states, implying that individuals with different time preferences for will assign different TTO values for the same health states. Because age and gender groups may differ in their time preference, they may also vary in assigning TTO values for a given health state. The TTO models must therefore be tested for age and gender differences.

TTO-1 allows for differences in the constant term (i.e. each individual is given a different constant term). This model is estimated for the Clinic and University samples. Joint Wald tests indicate that the average constants for males and females are different

for the Clinic sample at the 1% level of significance, and for the University sample at the 5% level of significance. Because of the small size of each sample and the fact that the differences between the average constants for males and females are in the opposite direction for the Clinic and University sample, it is difficult to draw any inferences from these differences.

The coefficients for the severity level of symptoms (i.e. $\alpha_1$ through $\alpha_{21}$) are tested to see whether they are different for males and females using the same methodology as in the previous chapter. At the 5% level of significance, the 21 coefficients for the symptoms severity levels are not jointly or individually different for males and females in both the Clinic and University samples. This indicates that males and females interpret changes in severity levels in the same manner. As was found for the SG, age is not a significant determinant in the TTO valuation of health states.

The initial set of parameter estimates for TTO-1, using the Clinic and University samples separately, are shown in table 6.

Table 6: TTO regression equations for Clinic and University samples

| | CLINIC SAMPLE | | UNIVERSITY SAMPLE | |
|---|---|---|---|---|
| **VARIABLE*** | **ESTIMATED COEFFICIENT** | **T RATIO** | **ESTIMATED COEFFICIENT** | **T RATIO** |
| $Hairloss_1$ | 0.0188 | 0.45 | 0.0365 | 1.04 |
| $Hairloss_2$ | 0.0123 | 0.26 | 0.0158 | 0.43 |
| $Hairloss_3$ | 0.0249 | 0.61 | -0.0037 | -0.11 |
| $Vomiting_1$ | 0.0654 | 1.91 | 0.1694 | 5.07 |
| $Vomiting_2$ | 0.0903 | 2.17 | 0.0916 | 2.81 |
| $Vomiting_3$ | 0.0206 | 0.49 | 0.0219 | -0.64 |
| $Depressed_1$ | 0.0487 | 1.23 | 0.1216 | 3.51 |
| $Depressed_2$ | 0.0293 | 0.79 | 0.0934 | 2.54 |
| $Depressed_3$ | -0.0168 | -0.92 | 0.0923 | 2.52 |
| $Breathing_1$ | 0.0733 | 1.61 | 0.0201 | 0.61 |
| $Breathing_2$ | 0.0459 | 1.07 | 0.0191 | 0.34 |
| $Breathing_3$ | 0.0817 | 1.83 | -0.0360 | -1.11 |
| $Swallowing_1$ | 0.0925 | 2.41 | 0.0145 | 0.44 |
| $Swallowing_2$ | 0.0610 | 1.44 | -0.0051 | -0.15 |
| $Swallowing_3$ | 0.0634 | 1.66 | -0.0066 | -0.19 |
| Chest $Pain_1$ | 0.1070 | 2.45 | 0.0934 | 2.72 |
| Chest $Pain_2$ | 0.0765 | 1.84 | 0.0574 | 1.54 |
| Chest $Pain_3$ | 0.0660 | 0.13 | 0.0741 | 2.23 |
| Coughing $Blood_1$ | 0.0906 | 2.25 | 0.1398 | 3.90 |
| Coughing $Blood_2$ | 0.0908 | 2.11 | 0.1794 | 4.77 |
| Coughing $Blood_3$ | 0.0589 | 1.36 | 0.0124 | 0.39 |
| Average Constant** | 0.2553 | | 0.2600 | |
| Male Constant** | 0.1918 | | 0.2820 | |
| Female Constant** | 0.2765 | | 0.2380 | |
| $\overline{R}^2$ | 0.7394 | | 0.5804 | |
| Deg of Freedom | 80 | | 129 | |

* For $Symptom_j$, the severity level is "not at all" for j=1, "a little" for j=2, "quite a bit" for j=3, while the base case is "very much".

** The constants are the average constants for the individuals in each group.

## 6.4 Comparison of University and Clinic samples

As in the SG analysis, the TTO regressions for the Clinic and University samples can be compared to determine whether there is a difference in the way the two samples assign TTO values. This comparison is performed to determine whether medical knowledge of lung cancer health states is an important factor in assigning TTO values.

Table 6 shows the estimated regression results for TTO-1, for the Clinic and University samples. The statistical tests behind the comparison of the two samples are performed using the same methodology as in the previous chapter. The average constants for females are not different for the Clinic and University samples at the 10% level of significance. The average male constants are different for the two samples at the 1% level of significance. Clinic males give lower TTO values for the health states. Unfortunately, not too much can be inferred from this result because of the small samples involved in the comparison (i.e. there are only 15 males in the University sample and 5 males in the Clinic sample). If the Clinic and student male constants are aggregated the average male and female constants are not significantly different at the 10% level of significance.

In a joint test the set of 21 coefficients for the symptom severity levels are not different across the two equations at the 5% level of significance. Individually, the coefficients for $\text{Vomiting}_1$ and $\text{Breathing}_3$ are the only two symptom coefficients that are different for the two samples at the 5% level of significance. The University sample has

a much large coefficient for $vomiting_1$ and a much lower coefficient for $Breathing_3$ than the Clinic sample. The University students interpret vomiting as a much more important symptom and breathing as a less important symptom. This difference in the way lung cancer symptoms are interpreted indicates that medical knowledge of the health states involved is an important factor in the way TTO values are generated. The implication of this result for the application of CUA is that care should be taken when choosing a sample for generating TTO values for lung cancer health states when vomiting or breathing changes are involved as a result of treatment.

The results for the model estimated with the combined Clinic and University sample are shown in table 7. The Vomiting and Breathing symptoms are allowed to vary between the two samples because of the statistically significant difference between their estimated coefficients.

Table 7: TTO regression for combined sample

| | COMBINED SAMPLE | |
|---|---|---|
| **VARIABLE*** | **ESTIMATED COEFFICIENT** | **T RATIO** |
| $Hairloss_1$ | 0.0285 | 1.07 |
| $Hairloss_2$ | 0.0232 | 0.83 |
| $Hairloss_3$ | 0.0096 | 0.38 |
| $Vomiting(student)_1$ | 0.1707 | 5.36 |
| $Vomiting(student)_2$ | 0.0963 | 3.07 |
| $Vomiting(student)_3$ | 0.0258 | 0.79 |
| $Vomiting(clinic)_1$ | 0.0586 | 1.69 |
| $Vomiting(clinic)_2$ | 0.0839 | 1.98 |
| $Vomiting(clinic)_3$ | 0.0231 | 0.56 |
| $Depressed_1$ | 0.0973 | 3.78 |
| $Depressed_2$ | 0.0656 | 2.52 |
| $Depressed_3$ | 0.0563 | 2.11 |
| $Breathing(student)_1$ | 0.0215 | 0.67 |
| $Breathing(student)_2$ | 0.0222 | 0.66 |
| $Breathing(student)_3$ | -0.0347 | -1.09 |
| $Breathing(clinic)_1$ | 0.0699 | 1.61 |
| $Breathing(clinic)_2$ | 0.0566 | 1.35 |
| $Breathing(clinic)_3$ | 0.0890 | 2.02 |
| $Swallowing_1$ | 0.0313 | 1.27 |
| $Swallowing_2$ | 0.0167 | 0.64 |
| $Swallowing_3$ | 0.0176 | 0.69 |
| $Chest\ Pain_1$ | 0.1118 | 4.23 |
| $Chest\ Pain_2$ | 0.0712 | 2.57 |
| $Chest\ Pain_3$ | 0.0599 | 2.31 |
| $Coughing\ Blood_1$ | 0.1238 | 4.67 |
| $Coughing\ Blood_2$ | 0.1384 | 4.94 |
| $Coughing\ Blood_3$ | 0.0219 | 0.88 |
| Average Constant** | 0.2529 | |
| Female Constant** | 0.2479 | |
| Male Constant** | 0.2603 | |
| Male Student Constant** | 0.2903 | |
| Male Clinic Constant** | 0.1706 | |
| $\overline{R}^2$ | 0.6523 | |
| Deg of Freedom | 224 | |

* For $Symptom_j$, the severity level is "not at all" for j=1, "a little" for j=2, "quite a bit" for j=3, while the base case is "very much".

** The constants are the average constants for the individuals in each group.

## 6.5 The preferred TTO regression equation

While the estimated regression equation in table 7 generally conforms to a priori expectations, there are some insignificant variables and unexpected coefficient patterns for the severity levels of some of the symptoms. The Hairloss and Swallowing variables can be omitted from the equation because none of the estimated coefficients for these variables are statistically significant at the 10% level. For the TTO technique, individuals are not interpreting the Hairloss or Swallowing symptoms as important factors when valuing health states.

The "quite a bit" severity level for Vomiting is not statistically significant, at the 10% level, for both the student and Clinic coefficients. The "quite a bit" variable for Vomiting can be omitted because individuals are not distinguishing between "quite a bit" and "very much" Vomiting when valuing health states. For the Clinic coefficients, the "not at all" coefficient is not different from the "a little" coefficient at the 10% level of significance, indicating that the "not at all" and "a little" severity levels for the Vomiting can be combined for the Clinic workers.

None of the three coefficients for the Breathing(student) variable is statistically significant at the 10% level. The students do not view Breathing as an important symptom. For the clinic workers there is no difference between the three Breathing coefficients at the 10% level of significance. The Clinic workers do not distinguish between "not at all", "a little", and "quite a bit" of trouble breathing when assigning TTO values.

For the Coughing Blood Symptom the coefficient for "not at all" and "a little" are not statistically different at the 10% level and the two severity levels can therefore be combined into one variable. The "quite a bit" coefficient is not statistically significant at the 10% level and can be omitted (i.e. combined with "very much").

The regression results incorporating these changes are presented in table 8.

Table 8: TTO regression results for the preferred model

| VARIABLE* | ESTIMATED COEFFICIENT | T RATIO |
|---|---|---|
| $\text{Vomiting(student)}_1$ | 0.1561 | 5.56 |
| $\text{Vomiting(student)}_2$ | 0.0814 | 2.93 |
| $\text{Vomiting(Clinic)}_{12}$ | 0.0660 | 2.36 |
| $\text{Depressed}_1$ | 0.1047 | 4.23 |
| $\text{Depressed}_2$ | 0.0716 | 2.93 |
| $\text{Depressed}_3$ | 0.0687 | 2.71 |
| $\text{Breathing(Clinic)}_{123}$ | 0.0778 | 2.27 |
| $\text{Chest Pain}_1$ | 0.1190 | 4.71 |
| $\text{Chest Pain}_2$ | 0.0757 | 2.96 |
| $\text{Chest Pain}_3$ | 0.0573 | 2.29 |
| $\text{Coughing Blood}_{12}$ | 0.1251 | 6.88 |
| Average Constant** | 0.2909 | |
| $\overline{R}^2$ | 0.6589 | |
| Deg of Freedom | 240 | |

* For $\text{Symptom}_j$, the severity level is "not at all" for $j=1$, "a little" for $j=2$, "quite a bit" for $j=3$, while the base case is "very much".

** The constant is the average constant for all individuals.

## 6.6 Interpreting the preferred TTO regression results

In this section the estimated regression equation in table 8 is examined to determine how individuals are assigning TTO values to health states. The estimated coefficients in the equation represent the change in TTO value that is due to a change in the severity level of that symptom from the base severity level "very much" to another severity level for the variable in question, holding the severity levels for all the other symptoms constant. These coefficients are examined to determine how the severity levels are interpreted for each of the symptoms, and to judge the relative importance of the symptoms in assigning TTO values to lung cancer health states.

For the Chest Pain and Depressed symptoms the coefficients for the three severity levels have the expected pattern. Specifically, as the severity level of Chest Pain or Depression gets better (i.e. moves from "very much" to "not at all") a higher TTO utility value is assigned to the health state. The Coughing Blood symptom is interpreted as either bad or good. For this symptom people are not distinguishing between the worst two severity levels (i.e. "very much" and "quite a bit") and are also not distinguishing between the best two severity levels (i.e. "not at all" and "a little"). The vomiting symptom is interpreted by the clinic workers in the same manner as the coughing blood symptom. The clinic workers interpret vomiting as either good (i.e. "not at all" or "a little") or bad (i.e. "very much" or "quite a bit"). For the Vomiting symptom students interpret the "quite a bit" and "very much" severity levels as equivalent, showing that they view quite a bit of vomiting to be as equally bad as very much vomiting. The

difference in the interpretation of the vomiting symptom for the two samples is that the students distinguish between the "not at all" and "a little" severity levels, while the clinic workers do not distinguish between these two severity levels. The Breathing symptom is also interpreted differently by the students and the clinic workers when assigning TTO values to health states. The students don't view the breathing symptom as important, while the clinic workers only view "very much" trouble breathing as important when assigning TTO values. The clinic workers don't distinguish between the "not at all", "a little", and "quite a bit" severity levels for the Breathing symptom. The Hairloss and Swallowing symptoms are not included because they are not important symptoms in assigning a TTO value for health states.

As explained in the previous chapter the relative magnitudes of the coefficients represent how important each symptom severity level is in assigning TTO values to health states. From table 8 Vomiting(student)$_1$ is the largest coefficient at .156, while Coughing Blood$_{12}$, Chest Pain$_1$, and Depression$_1$ have coefficients around .11, and Breathing(Clinic) and Vomiting(clinic) have smaller coefficients of around .07. These magnitudes are useful in evaluating treatment options in relation to the symptoms in a cost utility framework. The magnitude of the coefficients can be taken into account when considering how different treatment options affects symptoms. For example, if treatment using the same resources can either decrease the severity level of coughing blood from "very much" to "a little" or decrease the severity level of chest pain from "very much" to "a little", then the treatment should be administered for the coughing blood symptom because it increases the TTO utility value the most.

# CHAPTER 7: INDIRECT COMPARISON OF SG AND TTO

In chapter 4 it was noted that problems associated with the direct comparison of SG and TTO can be overcome by using an indirect comparison to test whether the SG and TTO are equivalent. The indirect comparison tests whether individuals evaluate the seven symptoms in the same manner for the two HSIs by comparing the models that were presented earlier for the SG and TTO techniques. If the SG and TTO are equivalent the coefficients should be the same whether the SG or TTO is the dependent variable. Even if the coefficients in the estimated regression equations are biased because of the limited dependent variables, both models will be biased in the same manner and tests of equivalence between the SG and TTO will not be affected.

To compare the SG and TTO models an identical model for each HSI is estimated and the coefficients for the two HSIs are compared using Wald tests. The model used in the comparison is C-1.

$$\begin{aligned} HSI = {} & \alpha_1 \text{ Hairloss}_1 + \alpha_2 \text{ Hairloss}_2 + \alpha_3 \text{ Hairloss}_3 + \alpha_4 \text{ Vomiting}_1 \\ & + \alpha_5 \text{ Vomiting}_2 + \alpha_6 \text{ Vomiting}_3 + \alpha_7 \text{ D*Vomiting}_1 \\ & + \alpha_8 \text{ D*Vomiting}_2 + \alpha_9 \text{ D*Vomiting}_3 + \alpha_{10} \text{ Depressed}_1 \\ & + \alpha_{11} \text{ Depressed}_2 + \alpha_{12} \text{ Depressed}_3 + \alpha_{13} \text{ D*Depressed}_1 \\ & + \alpha_{14} \text{ D*Depressed}_2 + \alpha_{15} \text{ D*Depressed}_3 + \alpha_{16} \text{ Breathing}_1 \\ & + \alpha_{17} \text{ Breathing}_2 + \alpha_{18} \text{ Breathing}_3 + \alpha_{19} \text{ D*Breathing}_1 \\ & + \alpha_{20} \text{ D*Breathing}_2 + \alpha_{21} \text{ D*Breathing}_3 + \alpha_{22} \text{ Swallowing}_1 \\ & + \alpha_{23} \text{ Swallowing}_2 + \alpha_{24} \text{ Swallowing}_3 + \alpha_{25} \text{ Chest Pain}_1 + \alpha_{26} \text{ Chest Pain}_2 \\ & + \alpha_{27} \text{ Chest Pain}_3 + \alpha_{28} \text{ Coughing Blood}_1 + \alpha_{29} \text{ Coughing Blood}_2 \\ & + \alpha_{30} \text{ Coughing Blood}_3 + \beta_1 \text{ IP}_1 + \ldots + \beta_k \text{ IP}_k + \varepsilon_5 \end{aligned} \qquad \text{(C-1)}$$

In this equation D is a dummy variable equal to 1 if the person is from the Clinic and zero otherwise. The rest of the variables in the equation are the same as presented earlier. This is the most flexible model for both HSIs and does not bias the results of the comparison by pre-specifying the way severity levels for symptoms are interpreted. The Breathing, Vomiting, and Depression variables are allowed to vary for the Clinic and Student samples because of the test results for the SG and TTO HSIs. The results for C-1 are presented in table 9 for both the SG and TTO equations. For ease of interpretation and comparison, the coefficients for Vomiting, Depressed, and Breathing are presented as the coefficients for each sample rather than as the coefficients for the interactive dummy variables shown in C-1.

The parameter estimates in table 9 can be used to compare the SG and TTO methods for valuing health states. However, statistical tests must be performed to determine whether the differences between the two sets of estimated coefficients are statistically significant. The statistical tests for comparison are performed by treating SG and TTO as one variable, estimating C-1, and then testing whether the coefficients are different for SG and TTO by using multiplicative and additive dummy variables[23].

[23] This is the same methodology used to test age, gender and sample differences in chapters 5 and 6.

Table 9: SG and TTO regressions for indirect comparison

| | SG | | TTO | |
|---|---|---|---|---|
| VARIABLE* | ESTIMATED COEFFICIENT | T RATIO | ESTIMATED COEFFICIENT | T RATIO |
| $Hairloss_1$ | 0.0515 | 1.97 | 0.0321 | 1.20 |
| $Hairloss_2$ | 0.0504 | 1.84 | 0.0273 | 0.97 |
| $Hairloss_3$ | 0.0261 | 1.04 | 0.0132 | 0.51 |
| $Vomiting(student)_1$ | 0.1562 | 4.96 | 0.1672 | 5.21 |
| $Vomiting(student)_2$ | 0.0839 | 2.71 | 0.0886 | 2.80 |
| $Vomiting(student)_3$ | -0.0046 | -0.14 | 0.0195 | 0.59 |
| $Vomiting(clinic)_1$ | 0.0601 | 1.76 | 0.0574 | 1.65 |
| $Vomiting(clinic)_2$ | 0.0592 | 1.42 | 0.0864 | 2.03 |
| $Vomiting(clinic)_3$ | 0.0531 | 1.29 | 0.0336 | 0.80 |
| $Depressed(student)_1$ | 0.1497 | 4.54 | 0.1307 | 3.89 |
| $Depressed(student)_2$ | 0.1750 | 5.10 | 0.0946 | 2.71 |
| $Depressed(student)_3$ | 0.0920 | 2.65 | 0.0945 | 2.67 |
| $Depressed(clinic)_1$ | 0.0196 | 0.50 | 0.0565 | 1.40 |
| $Depressed(clinic)_2$ | 0.0580 | 1.58 | 0.0365 | 0.98 |
| $Depressed(clinic)_3$ | 0.0337 | 0.85 | 0.0095 | 0.23 |
| $Breathing(student)_1$ | 0.0434 | 1.38 | 0.0212 | 0.66 |
| $Breathing(student)_2$ | 0.0469 | 1.42 | 0.0208 | 0.62 |
| $Breathing(student)_3$ | -0.0044 | -0.14 | -0.0338 | -1.07 |
| $Breathing(clinic)_1$ | 0.0771 | 1.81 | 0.0706 | 1.62 |
| $Breathing(clinic)_2$ | 0.0583 | 1.41 | 0.0462 | 1.09 |
| $Breathing(clinic)_3$ | 0.0951 | 2.19 | 0.0883 | 1.99 |
| $Swallowing_1$ | 0.0389 | 1.61 | 0.0338 | 1.37 |
| $Swallowing_2$ | 0.0288 | 1.14 | 0.0161 | 0.62 |
| $Swallowing_3$ | -0.0104 | -0.42 | 0.0171 | 0.68 |
| $Chest\ Pain_1$ | 0.1551 | 5.98 | 0.1138 | 4.29 |

| | SG | | TTO | |
|---|---|---|---|---|
| VARIABLE* | ESTIMATED COEFFICIENT | T RATIO | ESTIMATED COEFFICIENT | T RATIO |
| Chest $Pain_2$ | 0.0855 | 3.15 | 0.0695 | 2.51 |
| Chest $Pain_3$ | 0.0686 | 2.69 | 0.0587 | 2.26 |
| Coughing $Blood_1$ | 0.0941 | 3.60 | 0.1193 | 4.47 |
| Coughing $Blood_2$ | 0.1262 | 4.59 | 0.1347 | 4.80 |
| Coughing $Blood_3$ | 0.0384 | 1.56 | 0.02345 | 0.93 |
| Average Constant** | 0.2485 | | 0.2491 | |
| Female Constant** | 0.2881 | | 0.2489 | |
| Male Constant** | 0.1891 | | 0.2496 | |
| Female Student Constant** | 0.2544 | | 0.2221 | |
| Female Clinic Constant** | 0.3218 | | 0.2755 | |
| $\overline{R}^2$ | 0.7381 | | 0.7440 | |
| Deg of Freedom | 221 | | 221 | |

* For $Symptom_j$, the severity level is "not at all" for j=1, "a little" for j=2, "quite a bit" for j=3, while the base case is "very much".

** The constants are the average constants for the individuals in each group.

There are two ways in which C-1 can be different for SG and TTO: first, the coefficients for the symptom severity levels, $\alpha_1$, $\alpha_2$, ..., $\alpha_{30}$, can be different for SG and TTO; and second, the coefficients for the interpersonal variables, $ß_1$, $ß_2$, ..., $ß_k$, can be different for SG and TTO. These two types of differences are tested separately because of the different implications in terms of the equality of the SG and TTO HSIs.

The coefficients for the symptoms (i.e. $\alpha_1$, $\alpha_2$, ..., $\alpha_{30}$) are as a group not different for the SG and TTO models at the 10% level of significance. So as a group the symptom severity level coefficients are equivalent for the SG and TTO, but two of the

coefficients are individually different for the SG and TTO. The coefficients for Depression(student)$_2$ and Chest Pain$_1$ are the only coefficients for the symptoms that are different for the two models at the 5% level of significance. It can be seen in table 9 that the Depression(student)$_2$ and Chest Pain$_1$ are the most dissimilar coefficients in the two regression equations.

The symptom severity level coefficients, $\alpha_1$, $\alpha_2$, ..., $\alpha_{30}$, measure changes in HSI values due to changes in lung cancer health states. Changes in lung cancer health states have equivalent effects on SG and TTO except for two situations: first, changes from the "quite a bit" to the "very much" severity level of the Depression symptom in the student sample are valued differently by the SG and TTO methods; and second, changes from the "not at all" to "very much" severity level of Chest Pain are valued differently by the SG and TTO methods. Therefore changes in lung cancer health states are valued equivalently by SG and TTO if the two aforementioned changes in lung cancer health states are not evaluated.

The coefficients for the interpersonal variables (i.e. $ß_1$ through $ß_k$) must also be tested to see if they are equivalent for the SG and TTO. For CUA group averages of utility values for health states are used, so it is important to see whether the average constants for the different demographic groups are equivalent for the two HSIs. The only difference in the average constants for the two HSIs is that the SG varies over demographic groups while the TTO does not. At the 5% level of significance the SG average constants are different for males, student females, and clinic females, while there is no statistical difference between the demographic groups for the TTO average

constants. The average constant for the TTO is equivalent to the average female student SG constant, smaller than the average female clinic SG constant, and larger than the average male SG constant, at the 5% level of significance. This can be seen by examining the relative magnitudes of the average constants presented in table 9. A difference in the constant term means that the SG and TTO values for lung cancer health states are given different values on the scale. For example, males assign lower values to health states using the SG method than they do using the TTO method (i.e. because the average male constant is lower).

The discovery that the constant for the SG varies over demographic groups is an important result because it shows how studies that do not take demographic factors into account may yield biased results. For example, if this study only took a sample of student females the average constant would not be different for the SG and TTO and the false conclusion that the two HSIs are equivalent would be accepted. Also, if this study is performed without taking demographic factors into account the average constant for the TTO and SG would be found to be equal because the low male constant and high clinic student constant for the SG aggregate to be equivalent to the average TTO constant (i.e. it can be seen in table 9 that the average constants for the SG and TTO are almost identical). This difference in the average constant term implies that the TTO is a better practical HSI for use in CUA. This is because the TTO does not systematically vary over demographic groups. The SG varies over demographic groups with different risk attitudes, and is not very useful for interprogram evaluations because any difference in the risk attitudes of the samples could bias the results.

From this avalanche of results the question, "are the SG and TTO equivalent?", must be addressed. Unfortunately, the answer to this question is a definite "sometimes". That is to say, the SG and TTO are equivalent under certain conditions. The TTO and SG utility measures respond differently to changes in the severity level of Chest Pain from "very much" to "not at all", and Depression(student) from "very much" to "a little" differently. Therefore, changes in lung cancer health states are valued equivalently for the SG and TTO only if these two symptom severity levels are held constant.

# CHAPTER 8: CONCLUSION

## 8.1 Overview

In CUA the TTO technique for generating health state utility values is regarded as equivalent to the SG measure, even though there is no empirical justification in the literature to date. Because the SG measure is the only theoretically sound HSI, it is essential to know whether other HSIs used in CUA studies are equivalent to it. In this thesis the validity of using TTO as a proxy for SG is examined.

The equivalence of the SG and TTO is examined by using two different methods; the direct comparison, and the indirect comparison. The direct comparison is the traditional method for comparing HSIs. If the two HSIs are equivalent it is expected that when one is regressed on the other they are equal at every point. This can be tested by regressing the SG on the TTO and testing whether the intercept is equal to o and the slope is equal to 1. Unfortunately, having a limited dependent variable in the regression equation may bias the results towards rejecting the equivalence of the two HSIs.

The indirect comparison is used to overcome the uncertainty of having a limited dependent variable. The indirect comparison consists of building a model for each HSI and testing whether these models are equivalent for the SG and TTO. The models relate SG and TTO values to lung cancer health states, and also take into account the demographic variables of age, gender, and medical knowledge. If the SG and TTO are equivalent then the models relating HSI values to health states will also be equivalent.

Because any biases imposed by limited dependent variables will affect the models for both HSIs in the same manner, the comparison is not affected.

## 8.2 Summary of the results

The direct comparison of the SG and TTO indicates that they are not equivalent. The indirect comparison confirms this result. Although the two HSIs are not equivalent in the strictest sense, the indirect comparison does show that under certain conditions the SG and TTO are equivalent.

The direct comparison found that for both the Clinic and University samples the SG and TTO are not equivalent . This result agrees with Wolfson's (1982) findings in which he reports that the SG and TTO are not equivalent in a direct comparison. Unfortunately, these results can't be taken as conclusive because the dependent variable in the model used for the test is limited between 0 and 1.

The indirect comparison results concerning the equivalence of SG and TTO are mixed. Because this method compares the two models relating the SG and TTO generated utility values to lung cancer health states, it can be seen exactly how the HSIs are different. The models for each HSI consist of variables for the severity level of the symptom and interpersonal dummy variables, which allow individuals to vary in how they assign HSI values. The models can be different in two ways. First, the constant term (i.e. the base value assigned to lung cancer health states) can differ. Second, the coefficients for the severity level of the symptoms (i.e. the way in which changes in the

lung cancer symptoms are interpreted) can differ. The similarities between the two models are presented first followed by differences between the models.

As a group, the coefficients for the severity level of the symptoms are equivalent for the two models. Also if the average constant is used (i.e. the constants are not broken down into demographic groups) the constants for the two models are equivalent. From an overall perspective there is a lot of similarity between the SG and TTO models.

There are however two exceptions where SG and TTO are not equivalent HSIs. First, two of the symptom severity levels have coefficients that are different for the SG and TTO; "a little" Depressed (student sample only), and Chest Pain "not at all". For the student sample changes between "a little" Depression and "very much" Depression are valued differently for SG and TTO. Also changes between "very much" and "not at all" severity levels for Chest Pain are valued differently for the two HSIs. Changes in lung cancer health states have equivalent effects on SG and TTO except the two aforementioned cases.

The interpersonal constants represent the HSI value assigned to the worst lung cancer health state (i.e. all severity levels at "very much") for each individual. As a group the coefficients for the interpersonal variables are different for the SG and TTO models. This means that individuals are assigning different utility values to a given health state depending on whether the SG or TTO technique is used. In CUA, group averages are used to get values for health states. It is not important if individuals value the SG and TTO differently if the two HSIs are valued equivalently by the group as a whole. It is therefore important to examine the average values for the interpersonal variables. The

average constants for the SG and TTO are equivalent, but this result is misleading because the difference arises because the average SG constant is different over demographic groups while the TTO constant is not. The average SG male constant is smaller[24] than the average TTO constant, the average student female constant is not significantly different from the average TTO constant, and the average clinic female constant is larger than the average TTO constant. This means that males and Clinic females rate SG and TTO health states differently (i.e. lower and higher, respectively) while the student females rate the TTO equivalently except when the health states consist of the cases mentioned in the previous paragraph.

These results show that the SG and TTO are not equivalent in the strictest sense, but are equivalent under certain circumstances. Changes in SG and TTO values are equivalent only when the "a little" severity level for the depression symptom or the "not at all" severity level for chest pain are not included in the analysis. Therefore the two methods can be used interchangeably to measure changes in lung cancer health states when the two aforementioned symptom severity levels are not involved. Differences in the absolute value of health state utility values for the SG and TTO are found when the male or clinic female samples are used to generate values for health states. This difference in the constant term between the SG and TTO for some of the subgroups in the study is not a problem for CUA. It just means that one utility value is equal to the other by adding or subtracting a scalar (i.e. the difference between the two average constants. The TTO is a better practical utility measure for CUA because it is shown

[24] The statistical tests are presented in the previous chapter.

here to be much more stable over different demographic groups.

## 8.3 Conclusions and policy implications

The results of this study are based on the evaluation of lung cancer health states and are therefore specific to CUA in the treatment of lung cancer. However, some generalization to the rest of the health care domain are possible. First, due to the difference in the way the Depression and Chest Pain severity levels are interpreted the TTO can't always be used as a proxy for the SG. This difference between the two HSIs occurs in only 2 of the 30 severity levels for the lung cancer symptoms, so use of the TTO depends on the relative importance of these two symptom severity levels in defining health states. Therefore, there would be no problem using the TTO technique as long as changes from or to no chest pain and a little depression are irrelevant in the health states being examined.

Second, there are some factors in this analysis that show the TTO to be a better HSI than the SG. The SG values are different across the demographic groups with different risk attitudes. In this study, gender and medical knowledge are found to affect the way the SG values are assigned[25]. Hence, when collecting SG values for health states the sample must consist of either males or females separately or have an appropriate gender mix, from the viewpoint of the analysis. This means that for the SG

[25] Age is found not to be a significant factor, but the results are inconclusive because of the limited age distribution in the samples.

technique careful consideration must be taken to ensure that the sample used to generate utility values is representative of the desired population. Because the TTO is the same across sexes, or other demographic groups with different risk attitudes, it is a better HSI because it is not contaminated by risk aversion[26].

The difference in the way males and females assign SG values is a problem for CUA. The current thought in CUA is that demographic factors do not affect HSI values and many of the samples used for CUA analysis do not take into account these demographic factors or even use a representative sample. The results in this thesis indicate that SG generated utility values are influenced by these gender differences. Intra program comparison using the SG when these demographic factors are not taken into account is invalid. If males are consistently placing lower SG utility values on a given health state than females, and funds are allocated on the basis of need, as defined by lower utility values, then funds will be disproportionately allocated to the health states appraised by males.

In this study a difference is shown, for both the SG and TTO, between a sample of convenience (i.e. university students) and a sample with medical knowledge of the health states involved (i.e. the Clinic workers). It is hypothesized that these differences are due to health care professionals' familiarity with lung cancer symptoms. If it is knowledge of the health states that causes these differences, using individuals knowledgable about the health states for generating utility values may be important.

---

[26] Although the TTO technique can theoretically produce different values for individuals with different time preferences, there is no evidence of this occurring in this study.

The difference between the TTO and SG utility values makes it clear that decision makers must be very careful in taking utility values for health states from the literature to date. Because utility values in the literature are generated using either the SG, TTO or RS technique it would be inaccurate to blindly use published values[27]. One of the advantages of CUA is its ability to compare the results of different programs and treatments, but comparing QALY's that have been generated by different HSIs is like comparing apples and oranges.

This study shows the dilemma one faces when choosing a HSI for CUA, by identifying the strengths and weaknesses of the SG and TTO techniques. The SG is the theoretically correct measure to use but it incorporates risk aversion into its values, which means demographic groups with different risk aversion give different SG values. The TTO has no theoretical justification and while very close to the SG, is not completely equivalent. However the TTO unlike the SG is not contaminated by risk aversion.

This study touches on many areas of CUA that have yet to be fully examined. The areas that are in need of further research are; identifying the demographic influences on utility measures, and comparing the SG and TTO with a diverse set of health states so as to determine if and when the two HSIs are equivalent in a more general setting.

---

[27] Torrance (1986) provides a list of cost / QALY values for selected programs with the QALY values generated from a variety of different measurement techniques.

# REFERENCES

Aaronson, N., et al. (1987) "Multidimensional Approach to the Measurement of Quality of Life in Cancer Clinical Trials," in The Quality of Life of Cancer Patients, New York: Raven.

Aaronson, N., et al. (1991) "The EORTC Core Quality of Life Questionnaire, Interim Results of an International Field Study," in Effect of Cancer on Quality of Life, ed. D. Osaba, Boca Raton FL.: CRC Press, 185-203.

Allais, M., Hagen, O. (1979) Expected Utility Hypothesis and the Allais Paradox, Dordrech, Holland: Reidel.

Arrow, K. (1984) Collected Papers of Kenneth J. Arrow:Individual Choice Under Certainty and Uncertainty, Cambridge: Harvard.

Auld, C. (1991) "Preference Under Risk and the Standard Gamble Approach to Health Status Indexes," Honours Thesis at the University of Victoria.

Baird, J., Noma, E. (1978) Fundamentals of Scaling and Psychophysics, New York: Wiley.

Bush, J., Chen, M., Patrick, D. (1973) "Cost-Effectiveness Using a Health Status Index. Analysis of the New York State PKU Screening Program," in Berg, R. ed Health Status Indexes, Chicago: Hospital Research and Educational Trust.

Boyle, M., Torrance, G., Sinclair, J., Horwood, S. (1983) "Economic Evaluation of Neonatal Intensive Care of Very Low Birth Weight Infants," New England Journal of Medicine 308, 1330-1337.

Churchill, D., Torrance, G., Taylor, D. (1987) "Measurement of Quality of Life in End-Stage Renal Disease: the Time Trade-Off Approach," Clinical Investigative Medicine 10, 14-20.

Daboni, L., Montesano, A., Lines, M. (eds.) (1986) Recent Developments in the Foundations of Utility and Risk Theory, Dordrecht, Holland: Reidel.

Drummond, M. (1987) "Resource Allocation Decisions in Health Care: A Role for Quality of Life Assessments?," Journal of Chronic Disease 40, 605-616.

Drummond, M., Stoddard, G., Torrance, G. (1987) Methods for the Economic Evaluation of Health Care Programs, Oxford: Oxford Health.

Ellsberg, D. (1954) "Classic and Current Notions of Measurable Utility," Economic Journal 64, 528-556

Farquar, P. (1984) "Utility Assessment Methods," Management Science 30, 1283-1295.

Feeny, D., Torrance, G. (1989) "Incorporating Utility-Based Quality of Life Assessment Measures in Clinical Trials," Medical Care 27, S190-S204.

Fishburn, P. (1988) Nonlinear Preference and Utility Theory, London: John Hopkins.

Fisher, G. (1977) "Convergent Validation of Decomposed Multi-Attribute Utility Assessment Procedures for Risky and Riskless Decisions," Organizational Behaviour and Human Performance 18, 295-315.

Froberg, D., Kane, R. (1989) "Methodology for Measuring Health-State Preferences-II: Scaling Methods," Journal of Clinical Epidemiology 42, 459-471.

Goodwin, P., Feld, R., Evans, W., Pater, J. (1988) "Cost-Effectiveness of Cancer Chemotherapy: An Economic Evaluation of a Randomized Trial in Small-Cell Lung Cancer," Journal of Clinical Oncology 6, 1537-1547.

Green, W. (1990) Econometric Analysis, New York: MacMillan.

Grether, D., Plott, C. (1983) "Economic Theory of Choice and the Preference Reversal Phenomenon," American Economic Review 69, 623-738.

Hagen, O. (1979) "Towards a Positive Theory of Preferences Under Risk," in Expected Utility Hypotheses and the Allais Paradox ed. by M. Allais and O. Hagen, Dordrecht, Holland: Reidel.

Hershey, J., Schoemaker, P.(1985) "Probability versus Certainty Equivalence Methods in Utility Measurement: Are They Equivalent?" Management Science 35, 1213-1231.

Israel, L., Chahinian, A., (eds) (1976) Lung Cancer: Natural History, Prognosis, and Therapy, New York: Academic.

Jones-Lee, M. (1976) The Value of Life: an Economic Analysis, London: Robertson.

Kahneman, D., Tversky, A. (1979) "Prospect Theory: An Analysis of Decision Under Risk," Econometrica 47, 263-291.

Kaplan, R., Bush, J., Berry, C. (1978) "The Reliability, Stability, and Generalizability of a Health Status Index," American Statistical Association, 704-709.

Keeney, R., Raiffa, H. (1976) Decisions with Multiple Objectives: Preferences and Value Tradeoffs, New York: Wiley.

Lewellyn-Thomas, H., Sutherland, H., Tibshirani, R. (1982) " The Measurement of Patients' Values in Medicine," Medical Decision Making 2, 449-462.

Loomes, G., Sugden, R. (1982) "Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty," The Economic Journal 92, 805-824.

Luce, R., Raiffa, H. (1957) Games and Decisions, New York: Wiley.

McNeil, B., Weichselbaum, R., Pauker, S. (1980) "Speech and Survival: Tradeoffs Between Quality and Quantity of Life in Laryngeal Cancer," New England Journal of Medicine 305, 982-987.

Machina, M. (1987) "Choice Under Uncertainty: Problems Solved and Unsolved," Economic Perspectives 1, 121-154.

Machina, M. (1982) "'Expected Utility' Analysis Without the Independence Axiom," Econometrica 50, 277-323.

Mehrez, A., Gafni, A. (1989) "Health Years Equivalent: How to Measure them Using the Standard Gamble Approach," CHEPA Working Paper Series #22, McMaster U.

Mehrez, A., Gafni, A. (1989) "Quality Adjusted Life Years, Utility Theory, and Healthy-Years Equivalents," Medical Decision Making 9, 142-149.

Mehrez, A., Gafni, A. (1987) "The Optimal Treatment Strategy -- A Patient's Perspective," Management Science 33, 1602-1612.

O'Connor A., Boyd, N., Warde, D., Stolbach, L., Till, J. (1987) "Eliciting Preferences for Alternative Drug Therapies in Oncology Influence of Treatment Outcome Description," Journal of Chronic Disease 40, 811-818.

Ochs, J., Muhern, R., Kun, L. (1988) "Quality of Life Assessment in Cancer Patients," American Journal of Clinical Oncology 11, 415-421.

Page, A. (1968) Utility Theory: A Book of Readings, New York: Wiley.

Payne, J., Braunstein, M. (1971) "Preferences Among Gamble with Equal Underlying Probability Distributions," Journal of Experimental Psychology 87, 13-18.

Pliskin, J., Shepard, D., Weinstein, M. (1980) "Utility Functions for Life Years and Health Status," Operations Research 28, 206-224.

Quin, R. (1981) "The effect of Measurement Method on Preference Scales in a Medical Decision Making Context (abstr)," Medical Decision Making 1, 431.

Read, J., Quinn, R., Berwick, D., Fineberg, H., Weinstein, M. (1984) "Preferences for Health Outcomes: Comparison of Assessment Methods," Medical Decision Making 4, 315-329.

Sackett, D., Torrance, G. (1978) "The Utility of Different Health States as Perceived by the General Public," Journal of Chronic Disease 31, 697.

Schoemaker, P. (1980) Experiments on Decisions Under Risk: The Expected Utility Hypothesis, London: Martinus Nijhoff.

Schofield, J. (1987) Cost-Benefit Analysis in Urban and Regional Planning, London: Allen and Unwin.

Slovic, P., Lichtenstein, S. (1971) "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgement," Organizational Behaviour and Human Performance, 6, 649-744.

Slovic, P., Lichtenstein, S. (1983) "Preference Reversals: A Broader Perspective," American Economic Review, 73, 596-605.

Sullivan, D. (1966) Conceptual Problems in Developing an Index of Health, Vital and Heath Statistics, Series 2, Number 17, Public Health Services Publications No. 1000.

Torgerson, W. (1958) Theory and Methods of Scaling , New York: Wiley.

Torrance, G. (1971) A Generalized Cost-Effectiveness Model for the Evaluation of Health Programs. Doctoral Dissertation, Department of Industrial Engineering, State University of New York at Buffalo.

Torrance, G., Thomas, W., Sackett, D. (1972) "A Utility Maximization Model for the Evaluation of Health Care Programs," Health Services Research 7, 118-133.

Torrance, G. (1976) "Health Status Index Models: A Unified Mathematical View," Management Science 22, 990-1001.

Torrance, G. (1976) "Toward a Utility Theory Foundation for Health Status Index Models," Health Services Research 11, 349-369.

Torrance, G., Boyle, M., Horwood, S. (1982) "Application of Multi-Attribute Utility Theory to Measure Social Preferences for Health States," Operations Research 30, 1043-1068.

Torrance, G. (1986) "Measurement of Health state Utilities for Economic Appraisal," Journal of Health Economics 5, 1-30.

Torrance, G. (1987) "Utility Approach to Measuring Health-Related Quality of Life," Journal of Chronic Diseases 40, 593-600.

Tversky, A. (1975) "A Critique of Expected Utility Theory: Descriptive and Normative Considerations," Erkenntnis 9 163-173.

Varian, H. (1978) Microeconomic Analysis, New York: Norton.

von Neumann, J., Morgenstern, O. (1947) Theory of Games and Economic Behaviour, Princeton: Princeton.

Waner, K., Luce, B. (1982) Cost-Benefit and Cost-Effectiveness Analysis in Health Care, Ann Arbor, Michigan: Health Administration Press.

Weinstein, M., Stason, W. (1977) "Foundations of Cost-Effectiveness Analysis for Health and Medical Practices," New England Journal of Medicine 296, 716-721.

Whitmore, G. (1976) "The Mortality Component of Health Status Indexes," Health Services Research 11, 370-390.

Wolfson, A., Sinclair, A., Bombardier, C., McGeer, A. (1982) "Preference Measures for Functional Status in Stroke Patients: Interrater and Intertechnique Comparisons," in Kane, R. L., Kane, R. A., eds: Values and Long-Term Care, Lexington MA: Lexington Books.

# APPENDIX A: QUESTIONNAIRE

## LUNG COST EFFECTIVENESS STUDY QUESTIONNAIRE

**INTRODUCTION**
This voluntary questionnaire is intended to gather comparative information to be used with Dr. Coy's Lung Cost Effectiveness Study. All responses are strictly anonymous and confidential.

The basis of this questionnaire is to gather information about willingness to take different treatments when the risks and consequences are known for both the state of ill-health and the treatment. To do so, you will be asked to imagine you have the described state of ill-health (with known symptoms) which always has a life expectancy of twenty years. You will also be told about two treatment options and the associated consequences. Then you will be asked about your willingness to take the treatment.

Another way to describe the question, is to compare them to a weather report. At what point do you decide to take an umbrella or wear a raincoat? Do you do so if the forecast is 10% chance of rain? 40% chance of rain? 75% chance of rain? 100% chance of rain?

**QUESTIONNAIRE EXAMPLE**
**Scenario**
Imagine you are living in the described ill-health state with a known life expectancy of twenty years. You are offered an experimental drug which will have one of two results:

1. The drug, while not altering your life expectancy, will allow you to live your remaining twenty years in perfect health.

2. Upon taking the drug, you will die immediately.

Response I
Your willingness to take the drug will obviously depend on the probability of its success of producing result 1. If, for example, this probability was 100%, you would certainly take the drug. If, however, the probability of success was 70%; that is, there is a 70% chance you will get full health for twenty years and a 30% chance of instant death, you may or may not take the drug, depending on the severity of the ill-health state.

Your answer to scenario 1 should be the lowest probability of the drug's success for which you would still be willing to take the drug.

An answer of 65% would mean you would take the drug if its success rate was 65% or better but prefer not to take the drug (and consequently live twenty years in the condition) if the drug's success rate was lower than 65%.

Response II
Suppose the above drug did not exist . Instead, you are offered a different experimental drug which always gives you full health, but reduces your life expectancy to less than twenty years. Your response should reflect the minimum number of years in full health the drug would have to produce for you to be willing to take it. For example, a response of twelve would mean "I would require at least twelve years in full health over twenty years in the above state before I would be willing to take the drug".

**REMEMBER:**
You are free not to participate or to withdraw at any time and the results are anonymous and confidential.

**YOU ARE FREE NOT TO PARTICIPATE, OR TO WITHDRAW AT ANY TIME**

---

Please fill in the following:

| | | |
|---|---|---|
| [ ] Age | Male 0 Female 0 | Please enter staff code: [ ]<br>1. medical staff<br>2. nursing staff<br>3. technical staff<br>4. health records<br>5. other |

---

**INSTRUCTIONS**

Each of the following pages describes a state of ill health characterized by a number of symptoms and their respective severity. It is known that each ill-health state causes you to have a life expectancy of exactly twenty years. For each ill-health state, study the symptoms and their severity and try to imagine what it would be like to live in that state for the next twenty years. You are asked to give two responses for each health state.

**Response 1**

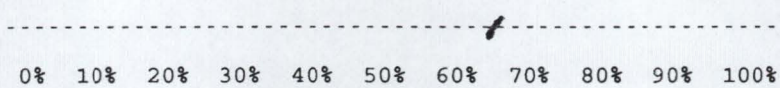
Imagine you are living in the described ill-health state. You are offered an experimental drug which would have one of two results:

1 The drug, while not altering your life expectancy, will allow you to live your remaining twenty years in perfect health.

2 Upon taking the drug, you will die immediately.

Your willingness to take the drug will obviously depend on the probability of its success, that is, of producing result 1. If, for example, this probability was 100%, you would certainly take the drug. If, however, the probability of success was 70%, that is, there is a 70% chance you will get full health for twenty years, and a 30% chance of instant death, you may not take the drug, depending on the severity of the ill-health state.

Your answer to response 1 should reflect lowest probability of the drug's success for which you would still be willing to take the drug. That is, if you would take the drug if the success rate was 65% or better, but prefer not to take the drug (and consequently live twenty years in the condition) if the drugs success rate was lower than 65%, you would clearly mark the probability line at 65%:




**Response 2**

Suppose the above drug did not exist. Instead, you are offered a different experimental drug which always gives you full health, but reduces your life expectancy to less than twenty years. your response should reflect the minimum number of years in full health the drug would have to produce for you to be willing to take it. For example, a response of twelve would mean "I would prefer to live twelve years in full health over twenty in the above state, but twenty years in the above state to eleven in full health."

| | Not at All | A Little | Quite a Bit | Very Much |
|---|---|---|---|---|
| Hair loss | | | ▲ | |
| Vomiting | ▲ | | | |
| Depression | | | | ▲ |
| Shortness of breath while resting | | ▲ | | |
| Trouble swallowing | | ▲ | | |
| Chest pain | ▲ | | | |
| Coughing blood | | | ▲ | |

RESPONSE 1

Mark the minimum probability of success the drug would have to have for you to be willing to take it. Remember: if it succeeds, it will produce full health for twenty years. If it fails, it produces immediate death.

---

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

RESPONSE 2

I would take the drug if it produced at least as many years in full health as circled below:

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

# VITA

Surname: Nield Given Names: James Andrew

Place of Birth: Winnipeg Date of Birth: April 7th, 1966

Educational Institutions Attended:

| | |
|---|---|
| University of Victoria | 1990 to 1992 |
| Bishop's University | 1987 to 1990 |

Degrees Awarded:

| | | |
|---|---|---|
| B.A. (Honours) | Bishop's University | 1990 |

Honours and Awards:

| | |
|---|---|
| University of Victoria Fellowship | 1991-92 |
| University of Victoria Teaching Award | 1991 |
| B.C. Sport Award | 1991 |
| Canadian Pacific Fellowship | 1990-91 |
| The Lord Strathcona Medal | 1990 |
| John W. Y. Smith Prize in Economics | 1990 |
| Chancellor's Scholarship | 1989-90 |
| Charles A. Shearson Memorial Prize | 1989 |
| University Prize in Economics | 1989 |

## PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies for such users or in response to a request from the Library of any other University, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying if this for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis: An Evaluation and Comparison of Utility Measures for Cost Utility Analysis

Author [redacted]

JAMES ANDREW NIELD

Dec 17, 1992