

USES AND ABUSES OF QALY ANALYSIS

By

Ann Michele Holmes

B. A. (Economics) University of Victoria

M. A. (Economics) Queen's University at Kingston

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
ECONOMICS

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December 1992

© Ann Michele Holmes, 1992

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of Economics

The University of British Columbia
Vancouver, Canada

Date 1 April 1993

Abstract

A major contribution of economics to health services research has been the development of QALYs (quality adjusted life years) as a measure of health status. This thesis investigates, in three essays, the use of QALYs in health care project evaluation and as an indicator of societal health.

The first essay examines the validity (defined as consistency with preferences) and feasibility of various QALY construction methods. Conditions for validity, derived from welfare principles, are used to assess the different methods. A new QALY instrument is devised that has interpersonal content (i.e. is valid for choices involving different individuals). Bias is shown to depend on various independence relationships within preferences. A number of these conditions are tested using data from the General Social Survey of 1985 (Canada. Statistics Canada [1987]).

The second essay examines the welfare properties of the QALY-based index as it is commonly employed to make health policy decisions. A comparison with alternative economic-based health indexes (human capital and willingness-to-pay) is provided. The QALY-based measure does indicate which treatment is best for an individual. In choosing patients for treatment, however, QALY-based measures probably discriminate against certain types of individuals, including those who are risk averse with respect to health and in poor health. In choosing between health programs, aggregate QALY-based measures do order community health profiles sensibly (except where people endure states worse than death), unlike the other measures considered. The QALY-based index may, however, favour unequal distributions of health.

The final essay assesses the appropriateness and feasibility of QALYs as a foundation for an index of societal health. Results suggest that, theoretically, the QALY serves as an imperfect measure of societal health, but that these problems are endemic to any index based on individual preferences. Using the best available data, a QALY based index is calculated to measure the level and distribution of ill-health in Canada and indicate where health policy can be most effectively targeted. The essay concludes with a discussion of what improvements in data collection are required to obtain more accurate figures.

Table of Contents

Abstract	ii
List of Tables	ix
List of Figures	xi
Acknowledgement	xii
1 Introduction	1
2 Separating Good Health Measures from Bad	8
2.1 Introduction	8
2.2 Part I: Validity	9
2.2.1 Motivation	9
2.2.2 Literature Review	10
2.2.3 Analysis	14
2.2.4 Independence Results	35
2.2.5 Equivalence Conditions	39
2.2.6 Instrument Selection	40
2.2.7 Summary of Validity Results	46
2.3 Part II: Implementation	47
2.3.1 Nature of the Problem	47
2.3.2 Literature Review	49
2.3.3 Analysis	54

2.3.4	Model	55
2.3.5	Production Method	56
2.3.6	Parameterization	64
2.3.7	Bias from Instruments	64
2.3.8	Bias from Aggregation	68
2.3.9	Synthesis	85
2.4	Conclusion	90
3	The Welfare Properties of Three Health Status Statistics	92
3.1	Introduction	92
3.1.1	Review of Health Status Measures	93
3.1.2	Evaluation Criteria	97
3.2	Theoretical Framework for Assessment	99
3.2.1	Individual's Optimization Problem	99
3.2.2	Derivation of the Health Status Indexes	104
3.2.3	Human Capital Measures (HK)	104
3.2.4	Willingness-to-Pay	106
3.2.5	QALY/HYE Measures	106
3.3	Choosing Treatments for a Given Individual	108
3.3.1	Exactness of the Health Status Index	108
3.3.2	Discussion	110
3.4	Choices Between Individuals	111
3.4.1	Variations Due to Endowments	112
3.4.2	Preference Variation	122
3.5	Choices Between Programs	124
3.5.1	Rationality of the Social Ordering	126

3.5.2	Welfarism	128
3.5.3	Ethics	131
3.5.4	Summary	133
3.6	Conclusion	134
4	A QALY Based Societal Health Statistic for Canada, 1985	136
4.1	Introduction	136
4.1.1	Purpose	136
4.1.2	Criteria for a Societal Health Index	137
4.1.3	Literature Review	140
4.2	Model	143
4.2.1	Theoretical Assessment	143
4.2.2	Completeness	145
4.2.3	Consistency	146
4.2.4	Ethical Content	150
4.2.5	Summary	153
4.3	Empirical Assessment	154
4.3.1	Data	154
4.3.2	Procedure	154
4.4	Results and Implications	160
4.4.1	Quality Adjusted Life Expectancy	160
4.4.2	Male-Female Differentials	163
4.4.3	The Importance of Morbidity	165
4.5	Avenues for Future Research	167
5	Conclusion	169

Bibliography	181
Appendices	190
A Chapter 2 Proofs	190
Appendices	214
B Simulation Results	214
Appendices	218
C Data Sources for Chapter 2	218
Appendices	224
D Empirical Results for Chapter 2	224
Appendices	237
E Proofs to Chapter 3	237
Appendices	241
F Proofs to Chapter 4	241
Appendices	250
G Empirical Analysis to Chapter 4	250
G.0.1 Data Description	250
G.0.2 Calculation of Satisfaction Estimates	253
G.0.3 Calculation of Expected Satisfaction	256

G.0.4	Linkage to QALY Values	258
G.0.5	Comment on Institutional Data	263
G.0.6	Calculation of QALY Averages	264
G.0.7	Life Expectancy	265
G.0.8	QALT Calculations	266

List of Tables

3.1	Conditions for Exactness	111
4.1	Quality Adjusted Life Expectancy, Canada and the Provinces	160
4.2	Quality Adjustment with Institutional Data, Canada and the Provinces	162
4.3	Male-Female Differentials, Canada	164
4.4	Morbidity by Major Category, Men	165
4.5	Morbidity by Major Category, Women	166
D.1	Estimation Results	224
D.2	Additive Tests	226
D.3	Joint Additive Tests	226
D.4	Multiplicative Tests (all data)	227
D.5	Joint Tests	227
D.6	Multiplicative Tests (< 30 yrs)	228
D.7	Joint Tests	228
D.8	Multiplicative Tests (30-65 yrs)	229
D.9	Joint Tests	229
D.10	Multiplicative Tests (> 65 yrs)	230
D.11	Joint Tests	230
D.12	Multilinear Tests	231
D.13	Joint Tests	232
D.14	Simulated Holistic QALY Values	233
D.15	Aggregated Values (cs)	233

D.16 Aggregated Values (sg)	234
D.17 Aggregated Values (tto)	234
D.18 Aggregated Values (es)	235
D.19 Aggregated Values (pe)	235
D.20 Distortions from cs Values	236
D.21 Distortions from sg Values	236
D.22 Distortions from tto Values	236
D.23 Distortions from es Values	236
D.24 Distortions from pe Values	236
G.1 Satisfaction Function Estimates (t-statistics in brackets)	255
G.2 Marginal Disutilities (taken at perfect health)	257

List of Figures

B.1	Time Trade-Off	214
B.2	Standard Gamble and Person Equivalents	215
B.3	Extended Sympathy	216
B.4	Most Likely Case	217

Acknowledgement

I would like to thank the members of my supervisory committee, David Donaldson, John Cragg, and Erwin Diewert, for their intellectual and emotional support during the preparation of this thesis. Despite the fact that none works in health economics, or perhaps because of this, they were able to guide my research on this subject in innovative directions and the final product has been greatly improved by their efforts. I would also like to thank my peers in “640” for their constructive comments and fellow inmates of C-block for their support.

This research was supported in part by the National Health Research and Development Program through National Health Fellowship number 6610-1557-47.

Chapter 1

Introduction

One of the most significant contributions economics can make to health services research is to provide appropriate means to evaluate the health care system so that policy may be directed to best serve the interests of society. These interests include not only the overall level of health and its distribution across members of society, but also those opportunities forgone because of the consumption of resources by the health care system.

Unfortunately, valuation techniques which are based on demand curves as revealed by market transactions (like cost-benefit analysis) are inappropriate because the link between the value of health outcomes and market data is made tenuous by peculiarities inherent in the health care system. Because health is not directly exchangeable between agents (with the possible exception of organ transplants), hedonic prices for health must be inferred from the demand functions for tradeable goods with health consequences. The legitimacy of such methods requires that markets for these goods exist without externalities (e.g. asymmetric information across patient types does not drive markets out of existence, insurance does not obscure the relationship between price paid and value received), and that consumers are willing and able to choose their purchases to maximize their own well-being (i.e. they understand completely the relationship between consumption of a good and its effect on health – which implies the health care professional’s role as provider of this information is redundant – and there is no habit formation or addiction that can cause deviations from this optimal

choice path over time). Since these characterizations are grossly atypical of actual health care markets, the valuations derived by such methods are usually invalid and can misdirect policy.

It is these same peculiarities of health care and related markets that ensure the failure of the price mechanism to allocate resources efficiently. This provides the justification for market intervention by a social planner. As demands on the health care system increase relative to the resources available to this sector (this due to the joint effect of rapid technological innovation allowing suppliers to offer more services, some of dubious worth, coupled with rising expectations on the part of consumers as to what these services can provide), decisions about which health care interventions to offer are apt to increase in frequency and importance.

The effectiveness of the social planner's directives depends on his or her ability to identify optimal health policies. Despite the difficulties noted above, a substantial amount of the evaluation literature has employed market data to evaluate health effects (see, for instance, Longmore and Rehahn [1975], Jones-Lee [1976], Mooney [1977]). While such methods are well established in the economics literature, this is offset by the measurement errors that are engendered by imperfect markets. An increasingly popular alternative is to value the health effects by means independent of the market and its distortions. Such a procedure requires, in effect, the construction of a health index. During the 1970's, a variety of such indexes were proposed (see McDowell and Newell [1987] for a compendium). Among these was the QALY (quality adjusted life year) index (Torrance et al. [1972]).

This thesis is an assessment of the QALY index as an instrument of policy formation. The QALY index is an appropriate topic of study for a variety of reasons. First, of all the health status indexes currently available, it has the strongest welfare foundation, being derived from preferences for health states. In fact, Torrance

(1976c) has demonstrated that most other health status indexes are approximations of the QALY index. Second, the QALY is a very flexible instrument and can be used to evaluate a wide range of health states in a variety of applications. Other indexes are restricted to specific contexts which are special cases of the QALY's. Third, and perhaps most important, the QALY is being increasingly employed by policy makers in practice. These applications include not only choosing the optimal treatment path for specific individuals (e.g. McNeil et al. [1981]), but also choosing between people for treatment (e.g. Boyle et al. [1983]), and choosing among projects for funding (e.g. Oregon Medicaid experiment as described in Klevit et al. [1991]). Furthermore, in some jurisdictions (e.g. Ontario, Australia), the index is on the verge of being institutionalized in the decision making apparatus. In their eagerness to find some systematic means to direct health policy that is free of market distortions, decision makers have employed QALYs in ever broader contexts where their validity is dubious. Research has focused on deriving a more comprehensive list of QALY values, leaving the analysis of the appropriateness of decisions made with these data outpaced by the applications which use them. This thesis attempts to narrow this gap between theory and practice.

As originally conceived, the QALY value acts as an adjustment factor on time alive. These scaling factors reflect the morbidity (or sickness) endured during a given period of time. Summing these values over all years of life yields a health index that incorporates both the morbidity and mortality (quality and quantity) aspects of health.

What distinguishes the QALY index from other health status indexes is not that length of life is scaled to reflect morbidity, but that these scaling factors are chosen to reflect the preferences of individuals for these different morbid states, assigning higher values to more preferred states of health than less preferred states of health. This

is assured by the methods used to derive QALY values: Hypothetical health states are described in a survey questionnaire. Respondents report their preferences for these states. The strength of these preferences is measured with a QALY instrument, a device which compares preferences for morbid states against preferences for some other measurable variable (the “metric variable”).

Since these values are based on assessments of outcomes only (which respondents should be able to evaluate without assistance), not production processes (where the information gaps prevent agents from linking goods consumed to health outcomes), the index should be free of market distortions. Furthermore, because the index is based explicitly on preferences, it has a stronger welfare foundation than most other health status indexes available (it assigns the highest values to states that leave the individual best off according to his or her own assessment).

The outstanding feature of QALYs is this relationship with preferences. This allows the possibility that the QALY may be used in utility based welfare assessments, free of the ethical problems inherent in traditional cost-benefit analysis (Blackorby and Donaldson [1990]). As a result, the concept of a QALY health index used in this thesis is somewhat broader than the standard definition, and any index derived from the individual’s preference ordering over health states as described above is included in the assessment, not just those that fit the strict definition of a QALY as a factor used to scale time alive. Thus, indexes such as Mehrez and Gafni’s (1989) healthy year equivalent (where both the quality and quantity aspects of a health state are assessed together in a single function) are treated as special cases of QALY indexes.

Through the 1970’s, QALY research focused on the development of decision statistics (methods for combining QALY values with other relevant data and accompanying rules to determine policy priorities) for a wide variety of policy applications. These included: the quality adjusted lifetime (the sum of the QALY values over all years of

life), used to evaluate optimal treatment paths; the cost-utility ratio (the difference in quality adjusted lifetimes caused by a health care project divided by its cost), used to determine how the health care budget should be allocated across different projects; the societal health index (the mean QALY or mean quality adjusted lifetime), used to measure the health status of a community (Torrance et al. [1972], Torrance [1976a,b,c]). Although Torrance (1986) subsequently argued that QALYs were only intended as a measurement device for input-output analysis, it is clear from the literature that much of the appeal of QALY analysis is its (presumed) consistency with welfare maximization (i.e. that decisions based on QALYs lead to society being as well off as possible). This has become increasingly apparent as policy moves towards a “patient centred ethic” (i.e. consumer sovereignty).

By the 1980's, other researchers began to raise concerns about the ethical underpinnings of QALYs. These concerns included (i) whether the use of QALYs as scaling factors produced a health status index that reflected preferences over the whole domain of health (see Pliskin et al. [1980], and Mehrez and Gafni [1989]), (ii) the underlying interpersonal assumptions of QALY-based decision statistics (see Harris [1987], Hilden [1985]), (iii) the social ethics of aggregate QALY statistics (Loomes and McKenzie [1989]), and (iv) whether resource allocations based on QALY data were consistent with welfare maximization (see Anderson et al. [1986], Birch and Donaldson [1987], Feeny and Torrance [1989]). These assessments have been incomplete, covering only select QALY applications. Some have raised the possibility that welfare inconsistencies may arise, but not under what conditions (Loomes and McKenzie, Harris, Hilden), while others have used assessment criteria that are not utility based and therefore do not reveal if the statistics are consistent with welfare maximization (Anderson et al., Birch and Donaldson, Feeny and Torrance). Virtually none has provided superior alternatives or strategies to compensate for biases. This might explain

why such research has failed to curb the use of QALY data in decision-making.

The three essays of this thesis constitute an attempt to close the gap between the theory and practice of QALY based evaluations, and to evaluate the index as an indicator of health status for individuals and society and its suitability for directing resource allocation in the health sector. The assessment is utility based: the QALY or associated decision statistic is evaluated as to whether or not it can identify the state which the individual (or society) most desires. When currently employed statistics fail to do this, the probable distortions on decision making are identified and unbiased alternatives are discussed. Also considered is the feasibility of implementing the various alternatives in practice.

The first essay examines the validity (defined as consistency with preferences) and feasibility of various QALY construction methods. Necessary and sufficient conditions for validity are derived from welfare principles and these are used to assess the different methods. A new QALY instrument is devised that has interpersonal content (i.e. is valid for choices involving different individuals). The degree of bias is shown to depend on various independence relationships within preferences. A number of these conditions is tested using data from the General Social Survey of 1985 (Canada. Statistics Canada [1987]). The more important conclusions drawn are: (1) while there does not exist any one QALY instrument that is universally valid, for every current application using QALY data, there exists at least one instrument that is valid; (2) over broadly defined categories of morbidity, multiplicative aggregation methods can be used to reduce significantly the costs of QALY data collection without inducing bias.

The second essay examines the welfare properties of a QALY-based index when it is used as a health status measure in three types of allocation decisions commonly

faced by policy makers: choosing treatments for an individual, choosing an individual for treatment, and choosing programs for implementation. A comparison is made with the two other economic-based health status indexes, the human capital and willingness-to-pay measures. The QALY-based measure does indicate appropriate treatments for an individual, as does the willingness-to-pay measure. The human capital measure generally fails to do this. All measures discriminate against some types of individuals. The QALY-based measure likely discriminates against unhealthy people and people who are risk averse with respect to health, which distinguishes it from the other measures. As an aggregate health status statistic (used whenever broadly based programs are compared), the QALY-based measure orders community health profiles sensibly (unless states worse than death exist), unlike the alternative measures. Decisions made with the QALY-based index may favour unequal distributions of health.

The final essay assesses the appropriateness and feasibility of QALYs as a foundation for an index of societal health. Results suggest that, theoretically, the QALY serves as an imperfect measure of societal health, but that these problems are endemic to any index based on individual preferences. Using the best available data, a QALY based index is calculated to measure the level and distribution of ill-health in Canada and indicate where health policy can be targeted most effectively. The essay concludes with a discussion of what improvements in data collection are required to obtain more appropriate figures.

Chapter 2

Separating Good Health Measures from Bad

2.1 Introduction

In this chapter, the issue of how to obtain QALY (quality adjusted life year) values is addressed. Various methods are assessed for theoretical validity (i.e. whether they assign values to health states in a manner consistent with preferences) and ease of implementation (i.e. costs of obtaining a given set of values). Conditions for validity are shown to depend on the nature of the decision statistic used, where the choice of decision statistic depends on what relevant factors change between states of the world. It is shown that no single method of construction is valid in every QALY application. Conditions for each construction method to be valid in specific applications are derived. A new method is devised that, unlike those currently in use, is valid in situations involving interpersonal comparisons.

Implementation considerations are viewed as an optimization problem where the cost of obtaining a set of QALY values is minimized by choice of construction method subject to some acceptable level of bias (where bias is measured by the deviation from true values that results in incorrect policy choices being made). Two strategies are assessed: (1) substitution of QALY instruments (the method by which QALY values are derived from preferences), and (2) reconstruction of values for multiple illness states from the values for single illness states. Bias is shown to depend on independence in preferences over morbidity from non-morbid and other morbid factors

respectively. These independence conditions are evaluated using simulation analysis and empirical tests. Results suggest that the standard conceptual framework, where health consists of sub-groupings of morbidity characteristics, is appropriate, and that cost savings in QALY construction may be achieved without significant bias by the adoption of multiplicative aggregation structures. The optimal choice of additional cost saving measures depends on how the QALY is used in decision making.

This chapter is organized in two sections: issues of validity are addressed in the first, while implementation issues are addressed in the second.

2.2 Part I: Validity

2.2.1 Motivation

One of the main advantages of QALY indexes over other health status indexes is that they are derived from preferences. Because preferences for health states cannot be inferred from behaviour (since rational and informed choices between health states are seldom made or observed), they must be obtained in some experimental setting. A subject is presented with a range of health states and asked which he or she prefers. These preferences are then converted to a numerical scale by the use of a survey instrument, a device used to measure the strength of preference between two states. A variety of such instruments are available (those used in the past include category scaling (CS), magnitude estimation (ME), standard gamble (SG), time trade-off (TTO), and person equivalents (PE)). To these is added the extended sympathy time trade-off instrument (ES). These are discussed in greater detail below.

Much energy has been expended to find the “gold standard” of QALY instruments, one which is universally appropriate and against which all other QALY instruments are to be judged. Such a concept underlies all empirical studies comparing

the mapping functions produced by the different instruments. After twenty years, no consensus in this debate has been reached. This is partially because the literature has been so fragmented that assessments have often been based on partial evidence, but also because the whole issue of what is appropriate and when has not been adequately addressed.

In this paper, a rigorous definition of appropriateness (validity) is derived from utility theory and the various instruments are evaluated against the conditions generated by this definition. The analysis is both complete and uniform. Theoretical results are supported by examples of commonly used QALY based analyses.

Nature of the Problem

In essence, each QALY instrument generates a mapping function from the multidimensional health space to a single-dimensioned health index by comparing the value of various health states to some measurable entity (called the metric variable). Since the metrics and the means of comparison vary across instruments, one might expect the mapping functions to vary as well. This poses the problem of which instrument, if any, to use.

Recall that the purpose of QALY analysis is to assist with health care decision making. The thrust of this paper is that the QALY to be used is the one whose information, used in the appropriate decision statistic, leads to the correct policy choice being made. All subsequent analysis proceeds with this in mind.

2.2.2 Literature Review

Papers in this area can be found under the general rubric of validity studies. Validity has been approached from both theoretical and empirical perspectives.

Theoretical analyses have often been hampered by vague and ill-defined concepts of validity. The definition usually employed is that of content validity: “A measurement instrument is valid to the extent that it actually measures the phenomenon it claims to measure.” (Phillips as quoted in Torrance [1976b, p. 132]) Such a definition requires that the purpose of the measurement be clearly understood. Then all conditions for validity should be derived from this starting point. QALY values are used to evaluate states of the world according to preferences. It is inappropriate to say that the QALY must represent these preferences since the QALY alone usually does not provide enough information to evaluate all aspects of a state. Rather, it is typically combined with other relevant information in a decision statistic and it is this statistic which must represent preferences. The conditions for a QALY instrument to be valid must be derived in the appropriate context. Instead, many of the studies undertaken to date have investigated some particular characteristic without first justifying that such a property is either necessary or sufficient for valid measurement.

Like the theoretical approach, the empirical approach suffers from a poorly defined concept of validity. Until such time as a “gold standard” is identified, statistical differences must be taken in relation to an arbitrary point and are meaningless. There is no reason why convergence would be expected to occur around the correct value (e.g. all the instruments could be biased). However, once such a standard is established theoretically, such investigations will be very useful.

Theoretical Investigations Attention has been focused on two characteristics: whether the QALY instrument generates an interval scale (i.e. is cardinally measurable), and whether the mapping function depends on aspects of the survey instrument itself apart from the morbidity description. The justification for the former condition has been that the QALY is often employed to measure health status differences and

that the difference operator is only meaningful on scales with at least cardinal measurability. Such a position ignores the fact that differences usually occur over more than just morbidity (e.g. time alive may also differ between states) or those situations where only a ranking of levels, not differences, is adequate (e.g. measures of societal health, where the QALYs must be comparable, though not necessarily cardinal).

The independence condition is necessary if the non-morbid context in which the QALY is derived differs from the situation which it is supposed to evaluate. If preferences over morbidity are conditioned on these other factors, then a separate QALY value may have to be calculated for each context. Independence is not necessary for validity unless QALY values are used across contexts. In fact, if independence does not hold in preferences, it should not hold in QALYs either since these are supposed to be preference based.

Neither of these conditions can be considered necessary nor sufficient for validity in general. The appropriate set of conditions must be derived from the decision statistic. Furthermore, evaluation of these conditions has been done in isolation, as though either was sufficient for validity. Unless they are perfectly congruent, this cannot be the case.

Interval Properties: QALY results in this area have been based exclusively on the expected utility literature, notably Luce and Raiffa (1957) and Fishburn (1964). Torrance (1986) asserts that only the SG instrument generates a valid QALY function because, being based on choice under uncertainty, it recovers the cardinal utility function that orders differences as well as levels of morbidity. But this position ignores (1) recent evidence indicating the underlying axioms necessary for cardinality do not hold (see Shoemaker [1982] for a general discussion, Loomes and McKenzie [1989] for applications specific to QALY analysis) and (2) that other value functions, notably pseudo-money metrics (i.e. which are structurally equivalent to money metrics but

are based on different value units), like TTO and PE, also exist. Thus, work in this area is incomplete and biased in favour of SG. Questions of which value function, if any, is appropriate go unasked.

Context: Because the QALY is only a hypothetical construct, some concern exists as to whether or not it reflects actual choices. First, if preferences are defined only as a binary choice relationship, the QALY must be based on choice to have economic content (eliminating the CS and ME instruments)¹. A related concern is whether the QALY depends on aspects of the hypothetical state that do not concur with actual situations. Attention has been focused on the TTO instrument and its dependence on the survey time frame. Pliskin et al. (1980) establish the conditions for independence (the marginal utility of time for any given health level must be constant, conditions which may be overly strong for some QALY applications). Loomes and McKenzie (1989) review the ensuing 10 years work, most of which suggests this strong condition is not satisfied in general. They also describe implications for risk analysis (results very similar to Blackorby and Donaldson's [1988]). In response to this, Mehrez and Gafni (1989, 1991) modify the TTO and develop the healthy year equivalent (HYE), which is not dependent on time. Similar rigour has not been applied to the study of other instruments' dependence on survey variables. Nor have the implications of this dependence on decision making been studied.

Empirical Investigations The health services literature has focused not on the theoretical issues, but on whether there exist statistically significant differences between instruments². Studies by Stevens (1959), Torrance (1976b), Read et al. (1984),

¹There exists an alternative position, commonly held by utilitarians like van Praag (1968), that utility is a measure of satisfaction rather than a representation of the preference ordering (see Sen [1985] for a discussion). There would then be no basis for excluding CS or ME.

²Such analyses neglect those situations where the QALY need only rank states, not measure them.

Wolfson et al. (1982), and Rosser and Kind (1978) compare the empirical performance of CS, SG, and TTO. The general conclusion is that the instruments generate results that are highly correlated, but not equivalent, and that convergence appears to be situation dependent (some results conflict). No study covering all instruments with adequate research design has been attempted.

Avenues for Further Research

This review reveals there has been no complete and uniform analysis of validity across all instruments. Part of the problem appears to be a weak concept of validity. An appropriate set of sufficient conditions has yet to be developed, and the existing necessary conditions are often suspect because their derivation from the decision statistic is seldom apparent. Finally, the evaluation has been constrained to a set of well established instruments that were developed at least twenty years ago. With one exception (Mehrez and Gafni [1989, 1991]), no attempt has been made to find superior instruments outside this set.

2.2.3 Analysis

Purpose

The purpose of this section is to examine the validity of different QALY instruments. This is needed to establish which instrument generates the appropriate QALY values and to establish the cost-efficiency trade-offs across instruments.

Definitions

The above statement begs the question: what is validity? In this context, validity is defined as the ability of the QALY index to indicate the state most preferred by an

individual (or by society when decisions affect many individuals). Such a criterion will ultimately depend on how the QALY is applied in decision making. Some definitions useful in the construction of this concept are presented below.

equivalence describes what ranking properties two functions share:

- a) f and g , two functions, are ordinally equivalent (rank levels of x the same way) if and only if, for all x in their common domain, there exists an increasing monotonic function, ϕ , such that $f(x) = \phi(g(x)) \stackrel{o}{=} g(x)$.
- b) f and g are cardinally equivalent (rank differences in x the same way) if and only if, for all x in their common domain, $f(x) = ag(x) + b \stackrel{c}{=} g(x), a > 0$.
- c) f and g are ratio scale equivalent (rank proportions of x the same way) if and only if, for all x in their common domain, $f(x) = ag(x) \stackrel{r}{=} g(x), a > 0$.

uniqueness $f_y(x)$ (“the function f of x evaluated in situation y ”) is unique if and only if it generates one and only one value for any x for every possible y . Otherwise, $f(x)$ is not a function.

completeness f is complete if and only if the domain of f , $D(f)$, is the entire set of x (X).

measurability describes the classification of functions that are informationally equivalent (i.e. provide the same information about preferences):

- a) a function, f , is ordinally measurable if any $\{\tilde{f} | (\tilde{f}(x) = g(f(x)) \forall x \text{ in the domain})\}$, where g is some increasing monotonic function, is informationally equivalent to \tilde{f} . This set of functions is described by S^o .

The values generated by such functions cannot be manipulated by addition nor multiplication operations.

- b) a function, f , is cardinally measurable if any $\{\tilde{f} | (\tilde{f}(x) = af(x) + b, a > 0, \forall x \text{ in the domain})\}$ is informationally equivalent to f . This set of functions is described by S^c .

Such functions exhibit interval properties and are suitable for addition operations (i.e. the difference between two function values is meaningful).

- c) a function, f , is ratio scale measurable if any $\{\tilde{f} | (\tilde{f}(x) = af(x), a > 0; \forall x \text{ in the domain})\}$ is informationally equivalent to f . This set of functions is described by S^r .

Such functions are suitable for multiplicative and lower order operations.

independence $f(x; z)$ (“the function f of x conditioned on z ”) is independent of z if and only if $f(x; z) = f(x; z') \forall x, z$.

Assumptions

The following assumptions establish the framework for the subsequent analysis.

- a) Let ξ_k denote the k th element of the K -dimensional vector q of all morbidity characteristics, each with an upper and lower bound, denoted by $\bar{\xi}_k$ and $\underline{\xi}_k$ respectively. Then the set q is bounded by \bar{q} and \underline{q} . It is assumed that q may be described in a quantified fashion that is comprehensible to the survey respondent.
- b) Let x denote the m -dimensional set of all variables over which preferences are defined, including morbidity. This set includes q , and various “context” variables:

time alive in the state (t), other commodities (y), and personal characteristics (α). The last two (non-health) factors are grouped in the subset κ ($\kappa = (y, \alpha)$).

c) Let the survey respondent have preferences over gambles involving x , denoted by $\tilde{\mathfrak{R}}$. It is assumed these preferences are complete, reflexive, transitive, and continuous. Then assume there exists a utility function, $\tilde{U} : R^{ms} \rightarrow R^1$ (s denotes the number of possible states of the world), which represents these preferences in the sense $\tilde{U}(x'_{win}, p', x'_{loss}, 1 - p') \geq \tilde{U}(x_{win}, p, x_{loss}, 1 - p) \leftrightarrow (x'_{win}, p', x'_{loss}, 1 - p') \tilde{\mathfrak{R}} (x_{win}, p, x_{loss}, 1 - p)$, where p denotes the probability of receiving the “win” value of x , and $(1 - p)$ is the probability of receiving the “loss” value of x (this is easily extended to more than two states).

d) Assume the survey respondent has preferences over certain outcomes, denoted by \mathfrak{R} , and that these preferences may be represented by a utility function defined over certain outcomes, $U : R^m \rightarrow R^1$, in the sense $U(x') \geq U(x) \leftrightarrow x' \mathfrak{R} x$. In the case where U is homothetic in time, $U(x) \stackrel{\circ}{=} \mu(q, \kappa)t$ (where “ $\stackrel{\circ}{=}$ ” indicates the functions are ordinally equivalent). Assume further that preferences over gambles are separable from states of the world that never occur. Then the utility function over certain outcomes must also represent preferences over gambles when there is no uncertainty (i.e. $p = 1$): $U(x'_{win}) \geq U(x_{win}) \leftrightarrow x'_{win} \mathfrak{R} x_{win} \leftrightarrow (x'_{win}, 1, x'_{loss}, 0) \tilde{\mathfrak{R}} (x_{win}, 1, x_{loss}, 0)$. This implies $\tilde{U}(x_{win}, 1, x_{loss}, 0) \stackrel{\circ}{=} U(x_{win})$.

e) Assume preferences also exist over differences in outcomes, denoted by $\hat{\mathfrak{R}}$, and that these preferences are represented by a cardinal utility function, \hat{U} . If (x^1, x^0) denotes the move from x^0 to x^1 , then $\hat{U}(x^1) - \hat{U}(x^0) \geq \hat{U}(x^3) - \hat{U}(x^2) \leftrightarrow (x^1, x^0) \hat{\mathfrak{R}} (x^3, x^2)$.³

³Notice that the cardinal utility function is defined directly over outcome differences and not implicitly from expected utility over gambles. This concept of a value function holds regardless of

- f) Assume preferences vary across individuals according to some finite number of characteristics. Then it is possible to express the preferences, as represented in a utility function, of an individual as $U^i(q, t) = U(q, t, \kappa_i)$, where U^i is the utility function of individual i , and κ_i are the characteristics of person i which affect preferences for outcomes over (q, t) .
- g) Assume an individual has social preferences (\mathfrak{R}_W) defined over the outcomes endured by all N members of society ((x_1, \dots, x_N) , the subscripts denoting individuals), and that these social preferences may be represented by a social welfare function, W , in the sense that $W(x'_1, \dots, x'_N) \geq W(x_1, \dots, x_N) \leftrightarrow (x'_1, \dots, x'_N) \mathfrak{R}_W (x_1, \dots, x_N)$.
- h) Let $\varphi^j(q; t, \kappa)$ denote the QALY value for morbidity profile q , conditioned on the context (t, κ) , derived by instrument j ($j = \text{CS, ME, SG, TTO, ES, or PE}$). Let $\Gamma(\varphi^j(q; t, \kappa), z)$ denote the decision statistic which employs the QALY value used in the evaluation and z , non-morbid factors which are relevant to the decision. z may contain elements of x besides q that differ from those used in the QALY valuation (see below). To distinguish between these, those factors which condition the QALY function are denoted (t, κ) , while those used in the decision statistic are denoted $(\tilde{t}, \tilde{\kappa})$. The decision rule is to select the state with the highest value of Γ .

This last assumption requires some discussion. The QALY value, $\varphi^j(q; t, \kappa)$, measures the value of q as conditioned on (t, κ) . In a sense, it only provides information about the value of morbidity. In reality, projects seldom affect morbidity alone. The policy maker must base his or her decision on all effects of the project. To do this,

the validity of the von Neumann-Morgenstern axioms and is slightly different from the value function typically used in the decision sciences.

the QALY data must be incorporated with some measure of these other effects in a decision statistic, Γ . There have been three such decision statistics used in the past. First is the QALT (quality adjusted lifetime), used when both morbidity and mortality change as a result of the decision: $\Gamma(\varphi^j(q; t, \kappa), z) = \varphi^j(q; t, \kappa)\tilde{t}$ (this is used to compare different health states for any one individual or between individuals). In this case, z includes the actual time spent alive in the morbid state (\tilde{t}), which may differ from the hypothetical time frame on which the QALY value is based (the actual time alive may not be known at the time QALY values are calculated or may be so variable that the replication of the QALY valuation exercise for each possible survival curve is prohibitively expensive). Hence, z may contain elements of x besides q that differ from those used in the QALY valuation. Second is the CUR (cost-utility ratio), used to evaluate projects which affect health status at some level of costs: $\Gamma(\varphi^j(q; t, \kappa), z) = (\varphi^j(q^A; t, \kappa)\tilde{t}^A - \varphi^j(q^B; t, \kappa)\tilde{t}^B)/C$ (where A and B denote after and before the project respectively, and C denotes the cost of the project). In this case, the change in health is compared to the resources needed to achieve it. z includes not only the time spent in the state, but the costs of the project as well. The costs of the project are not included as context variables in the QALY valuation exercise because the patient does not usually incur these costs (because of health insurance), yet the decision maker is aware of the resource drain on society that results from this decision and wants to take this into account. Third is the *ex post* measure of societal health, the sum of the QALTs of the N members of the society, used to measure community health: $\Gamma(\varphi^j(q; t, \kappa), z) = \sum_{i=1}^N \varphi^j(q_i; t_i, \kappa_i)\tilde{t}_i$. Again, z contains elements other than x (the number of people in society) and elements of x that differ from those on which the QALY value is based (the lengths of life).

Validity

An index is valid if it measures what it is supposed to measure. Most assessments of QALYs to date have evaluated validity by comparing empirical QALY values to other established measures of health status (content validity). Validity is established if the QALY index assigns values to health states in a manner similar to these other indexes. But this approach assumes these other measures are valid, which effectively negates the need to construct QALY values.

In this chapter, a theoretical assessment of the validity of the QALY index is undertaken. The validity results obtained do not require validity of any other health index, but only that preferences for health states reflect the well-being associated with those health states. One of the main contributions of this chapter is to operationalize the above definition of validity so that it may be used in a theoretical assessment.

QALYs are supposed to measure morbid states. More specifically, they must assign values to morbid states in a manner consistent with preferences so that more preferred states are assigned higher values than less preferred states. Recall, however, that the QALY is seldom used alone to evaluate states, but is usually combined with other relevant information in a decision statistic. Thus, the conditions for validity require that the decision statistic, which is defined over the QALY, be an exact representation of preferences over the state:

$$\Gamma(\varphi^j(q; t, \kappa), z) \stackrel{\circ}{=} U(q, \tilde{t}, \tilde{\kappa}) = U(\tilde{x}) \forall \tilde{x} \in D(U) \quad (2.1)$$

(where “ $\stackrel{\circ}{=}$ ” means the two functions are ordinally equivalent). Obviously, certain assumptions about the form of Γ and the content of z must be made (the form of Γ must be consistent with preferences, z can only contain fixed levels of any variables not included in the utility function) for the above condition to hold. The focus of this chapter is not the appropriateness of Γ , but the validity of $\varphi(q; t, \kappa)$. The

above condition imposes two (necessary) conditions on $\varphi(q; t, \kappa)$ for validity: (1) representation and (2) independence.

Consider the case where only morbidity is allowed to vary between states of the world. In this case, the QALY alone is a sufficient statistic to evaluate states of the world. Then the validity condition holds if and only if the QALY function is an exact representation of preferences over all morbidity levels for any given configuration of other factors, i.e.

$$\varphi^j(q; t, \kappa) \stackrel{o}{=} U(q, t, \kappa) \forall q \in D(U) \quad (2.2)$$

for any (t, κ) . In cases where changes in morbidity, rather than morbidity levels, are to be assessed, the QALY values must reflect value differences and the conditions on the QALY function become more strict: the QALY must be related to the cardinal utility function, \hat{U} , which represents preferences for such changes:

$$\varphi^j(q; t, \kappa) \stackrel{c}{=} \hat{U}(q, t, \kappa) \forall q \in D(\hat{U}), \quad (2.3)$$

(where “ $\stackrel{c}{=}$ ” indicates the functions are cardinally equivalent). The first necessary condition then may be summarized: the QALY function is (i) an (ordinally or cardinally) exact representation of preferences over its domain, and that this representation is (ii) complete and (iii) unique.

Now consider a more likely scenario where both morbidity and other factors are allowed to change between states of the world. In this case, the “context” on which the QALY function is based is likely to differ from the contexts over which the decision statistic is defined (the costs of recalculating the QALY value for every given context is prohibitively expensive; the actual context, which is used in the decision statistic, may not be known when the QALY values are obtained). In this case, the QALY function must represent preferences over morbidity not only for any given context, but for every possible context. Thus, states of the world are appropriately ranked

(assuming the conditions on Γ and z hold) if and only if

$$\bar{\Gamma}(\varphi^j(q; t, \kappa), \tilde{t}, \tilde{\kappa}) \stackrel{\circ}{=} U(q, \tilde{t}, \tilde{\kappa}), \quad (2.4)$$

where the fixed z elements are suppressed in $\bar{\Gamma}$. Given the right hand side of this equation is independent of (t, κ) , so must be the left hand side. But this is the case if and only if the QALY function is independent of the survey context, or

$$\varphi^j(q; t, \kappa) = \tilde{\varphi}^j(q) \forall (t, \kappa). \quad (2.5)$$

The evaluation of instruments proceeds according to the following structure: (1) does the QALY represent ordinal and cardinal preferences over morbidity for a given context, (2) is the QALY independent of changes in context, and (3) is the decision statistic consistent with preferences when context is allowed to change? Preceding this is a discussion of the instruments to be evaluated.

Description of Instruments and Relationships to Preferences

The following discussion includes the five instruments used in the past (the category scale, magnitude estimation, standard gamble, time trade-off, and person equivalents) and a new instrument presented here as a superior alternative for decisions involving interpersonal comparisons (the extended sympathy time trade-off).

Category Scaling (CS) With CS, health states are valued directly relative to one another. The respondent is asked in the survey to place the described health state on a line in relation to two distinct identified points (usually of perfect health and a state equivalent to death) according to the relative worth of each. The distance of this point from the death mark, divided by the distance between the perfect health and death marks is the QALY value.⁴

⁴There is also a discrete version of CS. In this case, the respondent is presented with a fixed number of points (or categories) where the health states may be ranked, perfect health and death

Assuming value assignments are made according to preferences (since “values” are involved, one might be tempted to assume that some “true” (i.e. cardinal) utility function is used, although this assumption is not made here), then the respondent would report values, conditional on his or her context (described by (t, κ)), as $U(q, t, \kappa)$ = the value of the described health state, $U(q^1, t, \kappa)$ = the value of the better reference state (perfect health), $U(q^0, t, \kappa)$ = the value of the worse reference state (death equivalent), such that the QALY:

$$\varphi^{CS}(q; t, \kappa) = \frac{U(q, t, \kappa) - U(q^0, t, \kappa)}{U(q^1, t, \kappa) - U(q^0, t, \kappa)} \quad (2.6)$$

(the QALY function depends on the reference states as well, although this is suppressed in the notation since these states are assumed to be fixed).

CS requires that preferences reflect the level of satisfaction as well as the preference ordering, that this ordering exist, and that the best and worst states be closed.

Magnitude Estimation (ME) ME is the ratio scale version of CS (which is an interval scale). The researcher identifies a single reference state, q^1 , as a numeraire and the respondent reports the value of the described health state as a multiple (e.g. fraction) of the value of the reference state.

Assume, as before, that value assignments are made according to some utility function defined over outcomes. Then

$$\varphi^{ME}(q; t, \kappa) = \frac{U(q, t, \kappa)}{U(q^1, t, \kappa)}. \quad (2.7)$$

ME requires the same conditions on preferences as in the CS case (see above), in addition to the assumption that $U(q^1, t, \kappa) \neq 0$.

usually assumed to be in the extreme categories. This reduces respondent effort in evaluation since only approximations are required.

Standard Gamble (SG) The standard gamble is based on principles of indifference between games of chance and known intermediate outcomes, as described in Fishburn (1964). Under certain conditions, depending on preference axioms and how the problem is presented to the respondent, the instrument shows similarities to factionalization techniques in psychometrics.

The respondent is presented with a health state and is asked to pick the odds (subjective probabilities) for a game with fixed win-loss (q^1 and q^0 respectively) states (the standard or reference gamble) such that the respondent is indifferent between taking the gamble or the described health state (i.e. the respondent is asked to pick p and $1 - p$, given q , such that $\tilde{U}((q, t, \kappa), 1) = \tilde{U}((q^1, t, \kappa), p, (q^0, t, \kappa), (1 - p)))$. Then p becomes the numerical assignment of the ordering over health states:

$$\varphi^{SG}(q; t, \kappa) = p. \quad (2.8)$$

For p to exist, an ordering over gambles must exist and be continuous over q such that agents are willing to accept gambles involving risk of death; for p to be unique, monotonicity must exist in preferences. If the von Neumann-Morgenstern axioms hold, p may be solved for explicitly as a function of the cardinal utility function:

$$\varphi^{SG}(q; t, \kappa) = p = \frac{\hat{U}(q, t, \kappa) - \hat{U}(q^0, t, \kappa)}{\hat{U}(q^1, t, \kappa) - \hat{U}(q^0, t, \kappa)}. \quad (2.9)$$

Time Trade-Off (TTO) Like the SG, TTO is an economic instrument in that it involves decision making on the part of the respondent about how much of one good should be given up in order to receive more of another. The respondent is described some health state, q , and told it will last some (hypothetical and exogenous) number of time periods, t , whereupon death will occur. The individual is asked to pick the portion or proportion of this length of life $((t - m)/t)$ spent in perfect health (q^1) that is equivalent to the state initially described (i.e. $U(q, t, \kappa) = U(q^1, t - m, \kappa)$, the

hypothetical state is generally described as occurring with certainty). Then

$$\varphi^{TTO}(q; t, \kappa) = (t - m)/t. \quad (2.10)$$

Person Equivalents (PE) Like SG and TTO, PE is based on choice and has economic content, although it is based on ethical or social preferences rather than selfish preferences. In this case, the respondent is told that N people live in health state q and is asked how many persons can be given up (killed off or assigned health state q^0) if the rest are given perfect health (q^1) and leave society equally well off. Let W denote social preferences in the same way that U denotes selfish preferences. Then the problem becomes to find m such that $W(q_1, \dots, q_N, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) = W(q_1^1, \dots, q_{N-m}^1, q_{N-m+1}^0, \dots, q_N^0, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N)$. The QALY is

$$\varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) = (N - m)/N, \quad (2.11)$$

where the subscripts on the q s, t s, and κ s identify the individual affected by the health status change.⁵ The respondent must consider the values, needs, and prognoses of the N persons in the sample and, if they differ, the selection mechanism which decides who dies and who is cured. A description of the N individuals' characteristics must be included in the health state description.

Extended Sympathy Time Trade-off (ES) Sen (1985) has suggested that preference information obtained in the fashion of the first four instruments above is appropriate for intrapersonal valuations, but not for interpersonal valuations. While PE may be based on the ethical preferences necessary for such comparisons, its various administrative and cognitive difficulties, combined with the fact that individual characteristics necessary for such comparisons are seldom provided, make it unsuitable

⁵Health characteristic k endured by individual i is denoted $\xi_{k,i}$.

for interpersonal comparisons. Yet, most applications of QALY analysis require some interpersonal content in the values since they examine allocations across individuals.

One possible method that does incorporate interpersonal comparisons and may have greater acceptance among respondents than PE (because immediate death is not involved) is a numerical version of extended sympathy. As originally developed (see Arrow [1978]), extended sympathy involves choosing who is better off: person i in state a or person j in state b . Such exercises provide an ordering over health states and individuals so interpersonal comparisons are possible. In this form, the procedure lacks the necessary measurability properties for many QALY applications. This problem is overcome here by introducing time trade-offs across individuals' states (i.e. the respondent is asked to choose what amount of time in state b endured by person j leaves j as well off as i who endures state a for some specified time). The proportion of time j spends alive relative to i is then a measure of the relative advantage of i 's situation over j 's.

For instance, if one felt occupation affected preferences over health, one would change the standard time trade-off question to read: "John Doe currently has no use of his left hand and, as a result, works in a service sector job earning 10,000 dollars a year. He will live another ten years in this state. Bill Smith lives in perfect health and has a job in the manufacturing sector earning 20,000 dollars a year. How long will Bill Smith have to live to be as well off as John Doe?" The ES QALY value is the answer to this question divided by 10 years.

Because there are a finite number of personal characteristics, it is possible to define preferences over personal characteristics and outcomes (again, uncertainty does not enter the problem as posed) such that $U(q, t, \kappa_i) = \bar{U}^i(q, t) \forall i$. Then the problem is

to choose m such that $U(q_i, t_i, \kappa_i) = U(q_j, t_j - m, \kappa_j)$ where

$$\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) = (t_j - m)/t_i. \quad (2.12)$$

Representation Results

This section examines whether each of the above instruments assigns values to morbid states in a fashion consistent with preferences, either cardinal or ordinal, for any given context. The following lemmata are structured according to whether the QALY acts as an appropriate utility function for complete, reflexive and transitive preferences. Such a function must

- 1) over its domain preserve the preference ordering over morbid states, whether orderings are defined over levels (i.e. is the function ordinally equivalent to the utility function, U , which is defined over levels of x) or differences in morbidity (i.e. is the function cardinally equivalent to the cardinal utility function, \hat{U} , which is defined over changes in states),
- 2) have as its domain the entire set of morbid states over which preferences are defined,
- 3) have an image which varies only with relevant changes in preferences (e.g. if morbid states need only be ordered, then only changes in the preference ordering should affect the QALY value; if morbid state differences must be ordered, then changes in cardinal utility or strength of preferences should affect QALY values).

If (1) fails, then the QALY has no welfare content (a state with a higher QALY value is not necessarily preferred to one with a lower QALY value). If (2) fails, states beyond the QALY domain cannot be evaluated even though the respondent has preferences for such states. If (3) fails, care must be taken that extraneous factors

are held constant across QALY values for different states or the same health state may be assigned multiple values.

Lemma 1 (CS): Given the assumptions above pertaining to CS

- 1.a) over its domain, φ^{CS} preserves the ordering over outcomes for any given context (t, κ) .
- b) over its domain, φ^{CS} preserves the ordering over morbid state differences, for any given context, if and only if $U \in S^c(\hat{U})$.
- 2) the domain of φ^{CS} is the set of all q over which preferences are defined for given (t, κ) .⁶
- 3.a) φ^{CS} s based on different representations of the same preference ordering over outcomes (U and U') have identical images if and only if $U \stackrel{c}{=} U'$; otherwise, the functions are ordinally equivalent.
- b) φ^{CS} s based on different reference states $((q^1, q^0)$ and $(\bar{q}^1, \bar{q}^0))$ have identical images if and only if $U(q^1, t, \kappa) = U(\bar{q}^1, t, \kappa)$ and $U(q^0, t, \kappa) = U(\bar{q}^0, t, \kappa)$; otherwise, the functions are cardinally equivalent.

Proof: see Appendix A.

Lemma 1 (ME): Given the assumptions above pertaining to ME

- 1.a) over its domain, φ^{ME} preserves the ordering over outcomes for any given context (t, κ) if and only if $U(q^1, t, \kappa) > 0$.
- b) over its domain, φ^{ME} preserves the ordering over morbid state differences, for any given context and non-zero reference values, if and only if $U \in S^c(\hat{U})$.

⁶This condition imposes fixity of (t, κ) and therefore ignores situations where the domains of q and (t, κ) in preferences are not independent. Since this section assumes such fixity throughout, it does not seem appropriate to address issues of independence which assume variability.

- 2) the domain of φ^{ME} is the set of all q over which preferences are defined (for given (t, κ)) if and only if $U(q^1, t, \kappa) \neq 0$; otherwise, the function is undefined so the domain is the null set.
- 3.a) φ^{ME} s based on different representations of the same preference ordering over outcomes (U and U') have identical images if and only if $U \stackrel{r}{=} U'$; otherwise, the functions are ordinally equivalent.
- b) φ^{ME} s based on different reference states (q^1 and \bar{q}^1) have identical images if and only if $U(q^1, t, \kappa) = U(\bar{q}^1, t, \kappa)$; otherwise, the functions are ratio scale equivalent.

Proof: see Appendix A.

Lemma 1 (SG): Given the assumptions above pertaining to SG

- 1.a) over its domain, φ^{SG} preserves the ordering over outcomes for any given context (t, κ) .
- b) over its domain, φ^{SG} preserves the ordering over morbid state differences, for any given context, if and only if $\tilde{U}((q_{win}, t, \kappa), p, (q_{loss}, t, \kappa), 1 - p) \stackrel{o}{=} pU(q_{win}, t, \kappa) + (1 - p)U(q_{loss}, t, \kappa)$ (i.e. the von Neumann-Morgenstern axioms hold), where, by definition, $U \in S^c(\tilde{U})$.
- 2) the domain of φ^{SG} is the set of all q whose values are bounded by the values of the two reference states (q^1 and q^0) for any given level of (t, κ) .
- 3.a) φ^{SG} s based on different representations of the same preference ordering over gambles (\tilde{U} and \tilde{U}') have identical images. If the von Neumann-Morgenstern axioms hold, φ^{SG} s based on different representations of the same preference ordering over outcomes (U and U') have identical images if and only if $U \stackrel{r}{=} U'$; otherwise, the functions are ordinally equivalent.

- b) φ^{SG} s based on different reference states $((q^1, q^0)$ and (\bar{q}^1, \bar{q}^0)) have identical images if and only if $U(q^1, t, \kappa) = U(\bar{q}^1, t, \kappa)$ and $U(q^0, t, \kappa) = U(\bar{q}^0, t, \kappa)$; otherwise, the functions are cardinally equivalent if $\tilde{U} \stackrel{\circ}{=} \sum_i p_i U(q_i, t, \kappa)$, and are ordinally equivalent if not.

Proof: see Appendix A.

Lemma 1 (TTO): Given the assumptions above pertaining to TTO

- 1.a) over its domain, φ^{TTO} preserves the ordering over outcomes for any given context (t, κ) if and only if t is fixed; φ^{TTO} preserves the ordering over morbidity levels for any t if and only if $U(x) \stackrel{\circ}{=} \mu(q, \kappa)t$.
- b) over its domain, φ^{TTO} preserves the ordering over morbid state differences, for any given context, if and only if $U(q, t, \kappa) \stackrel{\circ}{=} \mu(q, \kappa)t$ (i.e. utility is homothetic in time) and $U \in S^c(\hat{U})$.
- 2) the domain of φ^{TTO} is the set of all q such that, for any given κ , $U(q^1, t, \kappa) \geq U(q, t, \kappa) \geq U(q^1, 0, \kappa)$ ⁷.

⁷Note: The domain of the time trade-off instruments may be extended if some means of borrowing time is introduced. Torrance (1982) suggests that respondents choose the proportion of time spent in perfect health as opposed to some state worse than death such that indifference is achieved between this state and never being born (i.e. $U(\cdot, 0, \kappa) = 0 = U(\bar{q}, (1-a)t, q, at, \kappa)$). While this value is related to the regular TTO index, the correction factor required to make the two consistent is a function of the surveyed state: Assuming homotheticity in time, $\varphi^{TTO}(q; t, \kappa) = 1 - a = \mu(q, \kappa)/(\mu(q, \kappa) - \mu(\bar{q}, \kappa))$. But normal TTO values are scaled to the numeraire $1/\mu(\bar{q}, \kappa)$, so to make this index comparable with the normal index one must scale by the factor $(\mu(q, \kappa) - \mu(\bar{q}, \kappa))/\mu(\bar{q}, \kappa) = f(q; \kappa)$. Since this correction factor is based on an unknown value, comparability cannot be achieved. If, instead, one told respondents that this particularly bad state was preceded by a set period of perfect health and that deductions were to be made over the entire period (i.e. $U(\bar{q}, (1-a)t, q, at, \kappa) = U(\bar{q}, t - m, \kappa)$), then $\varphi^{TTO}(q; t, \kappa) = (a\mu(q, \kappa)/\mu(\bar{q}, \kappa)) + ((1-a)\mu(\bar{q}, \kappa)/\mu(\bar{q}, \kappa))$. To make this comparable with the normal index, one would subtract $(1-a)$ and divide by a . Since a is known, this is a feasible operation.

- 3.a) φ^{TTO} s based on different representations of the same preference ordering over outcomes (U and U') have identical images. When utility is homothetic in time, φ^{TTO} s based on different representations of the same preference ordering over morbidity (μ and μ') have identical images if and only if $\mu \stackrel{r}{=} \mu'$; otherwise, the functions are ordinally equivalent.
- b) φ^{TTO} s based on different reference states (q^1) and (\bar{q}^1) have identical images if and only if $U(q^1, t, \kappa) = U(\bar{q}^1, t, \kappa)$; otherwise, the functions are ratio scale equivalent if utility is homothetic in time and are ordinally equivalent if not.

Proof: see Appendix A.

Lemma 1 (ES): Given the assumptions above pertaining to ES

- 1.a) over its domain, φ^{ES} preserves the ordering over outcomes for any given context $(t_i, \kappa_i, t_j, \kappa_j)$ if and only if the reference time frame (t_j) is fixed; φ^{ES} preserves the ordering over morbidity levels for any t_i if and only if $U(x_i) \stackrel{o}{=} \mu(q_i, \kappa_i)t_i$.
- b) over its domain, φ^{ES} preserves the ordering over morbid state differences, for any given context, if and only if utility is homothetic in time and $U \in S^c(\hat{U})$.
- 2) the domain of φ^{ES} is the set of all q_i such that, for any given (t_i, κ_i) , $U(q_j^1, t_j, \kappa_j) \geq U(q_i, t_i, \kappa_i) \geq U(q_j^1, 0, \kappa_j)$.
- 3.a) φ^{ES} s based on different representations of the same preference ordering over outcomes (U and U') have identical images. When utility is homothetic in time, φ^{ES} s based on different representations of the same preference ordering over morbidity (μ and μ') have identical images if and only

if $\mu \stackrel{r}{=} \mu'$; otherwise, the functions are ordinally equivalent.

- b) φ^{ES} s based on different reference morbid states (q_j^1) and (\bar{q}_j^1) have identical images if and only if $U(q_j^1, t_j, \kappa_j) = U(\bar{q}_j^1, t_j, \kappa_j)$; otherwise, the functions are ratio scale equivalent if utility is homothetic in time, and are ordinally equivalent if not.

φ^{ES} s based on different reference person types (κ_j and $\bar{\kappa}_j$) have identical images if and only if $U(q_j^1, t_j, \kappa_j) = U(q_j^1, t_j, \bar{\kappa}_j)$; otherwise, the functions are ratio scale equivalent if utility is homothetic in time, and are ordinally equivalent if not.

Proof: see Appendix A.

Lemma 1 (PE): Given the assumptions above pertaining to PE

- 1.a) over its domain, φ^{PE} preserves the social ordering over morbid states for any given context $(t_1, \dots, t_N, \kappa_1, \dots, \kappa_N)$ if and only if N , the number of people in the group under consideration, is fixed.
- b) over its domain, φ^{PE} preserves the ordering over outcomes for any given context if and only if (i) the social welfare function is welfarist, (ii) N is fixed or social welfare is homothetic, and (iii) $t_i = t, \kappa_i = \kappa \forall i$, or $\partial^2 U / \partial \xi \partial \xi_k = 0$ for all (ξ, ξ_k) and for all i , where ξ is an element of (t_i, κ_i) . φ^{PE} is consistent with the average preference ordering if and only if (i) the social welfare function is welfarist and homothetic and (ii) the selection rule is random and N is large.
- c) φ^{PE} preserves the ordering over morbid state differences, for given context, if and only if (i) the social welfare function is welfarist and homothetic, (ii) either $t_i = t, \kappa_i = \kappa \forall i$, or $\partial^2 U / \partial \xi \partial \xi_k = 0$, and (iii) $U \in S^c(\hat{U})$.

2) the domain of φ^{PE} is the set of all morbid status distributions where

$$\begin{aligned} & W(q_1^1, \dots, q_N^1, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) \\ & \geq W(q_1, \dots, q_N, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) \\ & \geq W(q_1^0, \dots, q_N^0, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) \end{aligned}$$

for any given distribution of (t_i, κ_i) .

3.a) φ^{PE} s based on different representations of the same social preference ordering (W and W') have identical images. If the social welfare function is welfarist and homothetic, then φ^{PE} s based on different representations of the same preference ordering over outcomes (U and U') have identical images if and only if $U' = aU + b^i \forall i = 1, \dots, N$.

b) Let S be the selection rule for deciding who is cured and who dies in the alternative state. Then φ^{PE} s based on different selection rules have identical images if and only if (i) the social welfare function is welfarist and homothetic and (ii) $U(q, t_i, \kappa_i) \stackrel{c}{=} v(q) + b^i(t_i, \kappa_i) \forall i = 1, \dots, N$.

c) φ^{PE} s based on different reference morbid states $((q^1, q^0)$ and (\bar{q}^1, \bar{q}^0)) have identical images if and only if

$$\begin{aligned} & W(q_1^1, \dots, q_{N-m}^1, q_{N-m+1}^0, \dots, q_N^0, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) \\ & = W(\bar{q}_1^1, \dots, \bar{q}_{N-m}^1, \bar{q}_{N-m+1}^0, \dots, \bar{q}_N^0, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N); \end{aligned}$$

otherwise, the functions are cardinally equivalent if the social welfare function is welfarist and homothetic, and ordinally equivalent if not.

Proof: see Appendix A.

Summary of Representation

These six lemmata demonstrate that assumptions are required for each of these instruments to represent legitimately preferences. To order morbidity levels, only SG requires no additional assumptions; CS and ME require assumptions be made that valuations be based on preferences, and the other instruments are exact only for the given levels of their respective metric variables (in the neighbourhood of the level of the metric, these QALY values linearly approximate the true ordering).

To order differences in morbidity levels, no instrument is exact without the imposition of assumptions: SG extracts the cardinal utility function only if the von Neumann-Morgenstern axioms hold, CS and ME require individuals to make valuations based on a “cardinal” utility function, and the others require linearity in the metric variable in the “cardinal” utility function (regardless of the representation of these preferences actually used in the exercise).

The completeness of the trade-off instruments depends on the choice of reference points, the values of which bound the states that may be evaluated. The feasibility of changing these reference points is discussed above and an improved method (which generates comparable values) is suggested. Changes in reference points change the interval to which the image of the QALY function is projected, thereby changing values. In those cases where the metric of valuation enters utility homothetically, the two images are cardinally related and may be spliced together with any two common observations. Otherwise, the two functions are very difficult to relate.

Finally, only SG is invariant to affine but not general transforms of the utility function⁸, so its image will change if and only if strength of preference changes. TTO,

⁸CS is as well, but, since this instrument is based on valuations that need not be cardinally related to the cardinal utility function, it may be that an affine change in the cardinal utility function may manifest itself as a non-affine change in the utility function used in the CS evaluation.

ES, and PE are invariant to any such transforms such that their images are perturbed only by changes in the ordering itself. If morbid states are to be ranked only, the latter situation is more desirable, while if morbidity differences are to be ranked, the former situation is more appropriate (such that only differences in preferences that matter affect the QALY values; otherwise, intransitivities can occur).

2.2.4 Independence Results

This section enlarges the policy scope of the QALY by assessing whether the QALY can still evaluate morbid states when the context changes. This is one of the conditions for validity when policy affects more than just morbidity. In addition, a major result from the previous section is that some instruments are dependent on the level of their respective metrics. Seemingly, the other instruments are independent of these conditioning factors (e.g. McKenzie and Loomes [1989] seem to imply the TTO is inferior to other instruments because of a dependency on time, but do not discuss whether other instruments are also dependent on time). This position ignores that all instruments are conditioned on context (t, κ) in the above analysis.

In this section, the conditions for each instrument to be independent of any non-morbid context variable are established. These results indicate when QALYs are valid measures of morbidity in different contexts. Context variables fall into two categories: those controlled by the researcher because they are specified in the health state description, and those that are supplied by the respondent by default. Dependence on the latter set of factors is more problematic because they are beyond the direct control of the researcher and can only be manipulated by sample selection of the respondent group.

Lemma 2 (CS) φ^{CS} is independent of any element of (t, κ) if and only if $U(x) \stackrel{c}{=}$

$v(q)w(t, \kappa) + z(t, \kappa)$ for all (q, t, κ) in the domain.

Proof: see Appendix A.

If not, differentiation of the expression for $\varphi^{CS}(q; t, \kappa)$ with respect to any element of (t, κ) , say \S , reveals $\frac{\partial \varphi^{CS}(q; t, \kappa)}{\partial \S} < / > 0 \leftrightarrow (U_{\S}(q, t, \kappa) - U_{\S}(q^0, t, \kappa)) - \varphi^{CS}(q; t, \kappa)(U_{\S}(q^1, t, \kappa) - U_{\S}(q^0, t, \kappa)) < / > 0$. A sufficient condition for the CS QALY value to be lower when the level of another good increases is that health status and this good be complements in the utility function.

Lemma 2 (ME) φ^{ME} is independent of any element of (t, κ) if and only if $U(x) \stackrel{r}{=} v(q)w(t, \kappa)$ for all (q, t, κ) in the domain.

Proof: see Appendix A.

If not, differentiation reveals $\frac{\partial \varphi^{ME}(q; t, \kappa)}{\partial \S} > / < 0 \leftrightarrow U_{\S}(q, t, \kappa) - \varphi^{ME}(q; t, \kappa)(U_{\S}(q^1, t, \kappa)) > / < 0$. The ME QALY value for a given morbidity level falls as the level of another good increases if health and this other good are substitutes in the utility function and the reference morbid state is preferred to the measured morbid state.

Lemma 2 (SG) If the von Neumann-Morgenstern axioms hold, and (t, κ) are fixed for all states of the world, then φ^{SG} is independent of any element of (t, κ) if and only if $\hat{U}(q, t, \kappa) \stackrel{e}{=} v(q)a(t, \kappa) + b(t, \kappa)$ for all (p, q, t, κ) in the domain.

Proof: see Appendix A.

If not, differentiation reveals that $\varphi^{SG}(q; t, \kappa)$ relates to \S in the same way as $\varphi^{CS}(q; t, \kappa)$.

Lemma 2 (TTO)

a. φ^{TTO} is independent of any element of κ if and only if $U(x) \stackrel{o}{=} U(v(q, t), \kappa)$

for all (q, t, κ) in the domain.

b. φ^{TTO} is independent of the level of t if and only if $U(q, t, \kappa) \stackrel{o}{=} t\mu(q, \kappa)$ for all (q, t, κ) in the domain.

Proof: see Appendix A.

If not, $\frac{\partial \varphi^{TTO}(q; t, \kappa)}{\partial s} > / < 0$ as $\frac{\partial MRS_{\xi, t}}{\partial s} < / > 0$ and $sign(\frac{\partial \varphi^{TTO}(q; t, \kappa)}{\partial t}) = sign(\frac{U_t(q, t, \kappa)}{U(q, t-m, \kappa)} - \varphi^{TTO}(q; t, \kappa))$. The TTO value increases with t if the marginal utility of time is higher in the measured state than in the reference state. The effect on the TTO value of any other factor change depends on how that factor affects the relative value of morbidity versus time.

Lemma 2 (ES)⁹

- a) If $\kappa_i \neq \kappa_j$, then φ^{ES} is independent of the level of any element of $\kappa_{i,j}$ if and only if $\frac{\partial U(x_i)}{\partial s} = 0$ for all (q, t_i, κ_i) in the domain.
- b) If $t_i \neq t_j$, then φ^{ES} is independent of the level of t_j , and the level of t_i , if and only if $U_t(x_i)t_i = U_t(x_j)(t_j - m)$ for all $(q, t_i, \kappa_i, t_j, \kappa_j)$ in the domain (i.e. U is homothetic in t).

Proof: see Appendix A.

If not, differentiation yields $\frac{\partial \varphi^{ES}(q, t_i, t_j, \kappa_i, \kappa_j)}{\partial s_i} > / < 0$ and $\frac{\partial \varphi^{ES}(q, t_i, t_j, \kappa_i, \kappa_j)}{\partial s_j} < / > 0$ as $\frac{\partial U}{\partial s} > / < 0$ and $\frac{\partial \varphi^{ES}(q, t_i, t_j, \kappa_i, \kappa_j)}{\partial t_i} > / < 0$ as $U_t(x_i)t_i > / < U_t(x_j)(t_j - m)$. Intuitively, the ES case is similar to the TTO case, although it is defined over a larger class of utility functions.

Lemma 2 (PE)

- a. φ^{PE} is independent of the level of (t, κ) (i.e. $t_i = t, \kappa_i = \kappa$) if and only if

$$W(q_1, \dots, q_N, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) \stackrel{o}{=}$$

⁹In this case, only situations with differing personal characteristics or time frames are considered; the converse situation reduces to the TTO case.

$$\tilde{W}(V(q_1, \dots, q_N), t_1, \dots, t_N, \kappa_1, \dots, \kappa_N)$$

for all (N, q_i, t_i, κ_i) in the domain. In the case where the social welfare function is welfarist, this requires (i) that social welfare be homothetic and (ii) $U(q_i, t_i, \kappa_i) \stackrel{c}{=} v(q_i)a(t_i, \kappa_i) + b(t_i, \kappa_i)$.

b. φ^{PE} is independent of the distribution of (t_i, κ_i) (i.e. $t_i \neq t_j, \kappa_i \neq \kappa_j$) if and only if

$$W(q_1, \dots, q_N, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) \stackrel{o}{=}$$

$$\tilde{W}(V(q_1, \dots, q_N), t_1, \dots, t_N, \kappa_1, \dots, \kappa_N)$$

for all (N, q_i, t_i, κ_i) in the domain. In the case where the social welfare function is welfarist, this requires (i) that social welfare be homothetic and (ii) $U(q_i, t_i, \kappa_i) \stackrel{c}{=} v(q_i) + b(t_i, \kappa_i)$.

Proof: see Appendix A.

As before, differentiation of the expression for PE reveals the direction of bias when these conditions do not hold. In the case where a non-metric variable changes (i.e. not the number of people alive), the effect on the PE QALY values depends on how this factor affects the willingness to trade off one person's health for another's (i.e. $\frac{\varphi^{PE}(q_1, \dots, q_N, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N)}{\partial \zeta} > / < 0 \leftrightarrow \frac{\partial MRS_{q_i, q_j}}{\partial \zeta} > / < 0$ where $\zeta \in (t_i, \kappa_i) \in [1, \dots, N]\}$).

Summary of Independence

These lemmata reveal an important distinction between the independence conditions for context variables set by respondents and those set by the researcher (the levels of the metrics). The latter are much more stringent since metrics must enter utility homothetically¹⁰, while preferences for morbid states need only be separable from

¹⁰Note that a function of one variable is homothetic if it is increasing in that one variable.

the other context variables (the exact form of this separability is determined by the class of transformations of utility that do not affect the QALY values for a particular instrument).

It is possible to determine how biased the QALY values are across contexts if the second derivative properties of the utility function are known. Independence requires null second derivatives: cross-context QALYs are essentially linear approximations to the true values. The degree and sign of the bias depend on the amount and direction of curvature in preferences.

2.2.5 Equivalence Conditions

Each QALY instrument is some transformation of some conditional utility function. The Representation section establishes the conditions necessary for the transformations to be the same (i.e. for the mapping to occur to the same interval), while the Independence section shows the conditions for the conditional orderings of morbidity implied by the QALY values to be the same regardless of the context in which the QALY is derived. Combining the results from the two sections, the following Proposition may be stated:

Proposition 1: Equivalence

- a) all instruments yield identical results, given appropriate choice of reference points, identical individuals who share a common but arbitrary level of non-health characteristics, and a homothetic, welfarist SWF, if and only if preferences may be represented by

$$\tilde{U}(x_s, p_s) \stackrel{r}{=} \sum_s p_s v(q_s) t_s w(\kappa_s),$$

- b) all economic instruments (SG, TTO, ES, PE) are identical, given the conditions

in (a), if and only if preferences may be represented by

$$\tilde{U}(x_s, p_s) \stackrel{c}{=} \sum_s p_s v(q_s) t_s a(\kappa_s) + b(\kappa_s),$$

c) both selfish economic instruments (SG, TTO) are identical, given appropriate choice of reference points, if and only if preferences may be represented by

$$\tilde{U}(x_s, p_s) \stackrel{c}{=} f\left(\sum_s p_s v(q_s) t_s, \kappa\right),$$

d) both interpersonal instruments (PE, ES) are identical, given the conditions in (a), if and only if preferences may be represented by

$$U(x) \stackrel{c}{=} v(q_i) t a(\kappa_i) + b^i(\kappa_i),$$

but if individuals are not identical, preferences must be of the form

$$U(x) \stackrel{c}{=} v(q_i) t + b^i(\kappa_i).$$

Proof: see Appendix A.

If, after appropriately rescaling values (by adjusting reference health states to map to the same interval), results differ, it can be concluded that these conditions do not hold. A choice must then be made between instruments.

2.2.6 Instrument Selection

This section serves two purposes: first, the implications of inexactness in the decision statistics are brought forward and, second, how to overcome problems of this nature by appropriately fixing the reference context is demonstrated.

In this section, four common policy spaces and their respective decision statistics are examined: (1) to choose between health states for a single individual (QALT), (2)

to measure the health status improvement from a project (CUR), (3) as a measure of societal health (sum of QALTs), and (4) to choose between patients for a given health status change (QALT). The conditions under which these decision statistics are exact are stated and their performance under realistic parameterizations of the utility function are considered.

QALY as an element in a health index The decision statistic in this case is $H(q, t) = \varphi^j(q)t$ (context variables are held fixed and suppressed in the notation). This health index orders (q, t) appropriately if and only if $U \stackrel{\circ}{=} v(q)t$. Suppose this is not so. Then from Lemma 1 (TTO), $\varphi^{TTO}(q)$ is inexact except when the survey and actual time frames coincide. The non-time based indexes order morbid states appropriately, but are inexact over sets of (q, t) . In fact, over (q, t) , only the TTO with time appropriately set is exact (this result can be found in Mehrez and Gafni [1989], although they do not examine the exactness properties of the other instruments).

Example: Let $U(q, t) = \int_0^t e^{-r\tau} v(q) d\tau$. This is a commonly accepted utility function defined over time. Usually, r is taken to be between .05 and .10. Then $\varphi^j(\cdot)t = v(q)t \neq U(q, t)$ unless $r = 0$ ($j=CS, ME, SG, \text{ and } PE$), whereas

$$\begin{aligned} t - m &= \varphi^{TTO}(q, t)t = \ln(1 - r(U(q, t)))^{-1/r} & \text{if } r > 0 \\ &= v(q)t & \text{if } r = 0 \\ &\stackrel{\circ}{=} U(q, t) & \forall r. \end{aligned}$$

The bias of using one of the other instruments may then be found: $B = (t - m[r = 0]) - (t - m[r > 0]) = (\frac{\partial(t-m)}{\partial r})dr = (\frac{U(q, t)}{1/r - U(q, t)})dr$. Since $dr \geq 0$, and it is assumed that $U(q, t) > 0$ in order for TTO to be well-defined, $B > 0$ if $1/U(q, t) > r$ and $B < 0$ if $1/U(q, t) < r$. But since $(t - m)$ is undefined for the latter condition, it can be asserted that B is always positive, more so the larger is t or r .

This implies that, under usual preferences, the TTO instrument assigns lower

values to morbid states than the other QALY instruments. If values derived by different instruments are compared, morbid states evaluated by non-TTO instruments are overvalued relative to those valued by TTO (morbid states are consistently ranked within any given instrument, but not when different instruments are used). The problem becomes significant when both morbidity and longevity change between states. Non-TTO instruments (inappropriately) undervalue states with greater length of life, whereas the TTO instrument assigns appropriate weight to length of life and morbidity during it. Thus, if non-TTO instruments are used, health care policy will improve the quality of life more and the length of life less than what would achieve the greatest level of satisfaction.

QALY as a measure of health status improvement The decision statistic in this case is $\varphi^j(q^A)t^A - \varphi^j(q^B)t^B$ (non-health factors are fixed and suppressed in the notation). This is exact if and only if $\varphi^j(q)t \stackrel{c}{=} \hat{U}(q, t)$ (i.e. the QALY value is cardinally related to the cardinal utility function so that differences in the QALY value appropriately measure changes in morbidity). From Lemma 1 (SG), if the von Neumann-Morgenstern axioms are imposed, $\varphi^{SG} \stackrel{c}{=} \hat{U}$ for any given context (i.e. the hypothetical and actual time frames coincide). Additional and even more unrealistic assumptions must be imposed for the other instruments to generate this result. However, as is shown below, this is not sufficient for the SG to rank health status differences appropriately because neither the dependence of SG on t nor the functional form of Γ have been considered. In fact, the very disappointing result is generated that, in general, the SG ranks only changes in morbidity appropriately.

Consider first the case where $\hat{U} = v(q)t$. Then $\varphi^{SG}(q) = av(q) + b$ and $\varphi^{TTO}(q) = v(q)$. The decision statistic values are:

$$\Delta\varphi^{SG}t = (av(q^A) + b)t^A - (av(q^B) + b)t^B, = a\Delta\hat{U} + b(t^A - t^B) \quad (2.13)$$

(where Δ is the difference operator), which is exact if either $b = 0$ or $t^A = t^B$ in all applications. Conversely, since TTO employs some utility function $U \doteq v(q)t$,

$$\Delta\varphi^{TTO}t = (av(q^A))t^A - (av(q^B))t^B = a\Delta\hat{U}, \quad (2.14)$$

which is exact regardless, suggesting TTO is superior to SG.

Now consider the case $\hat{U} = v(q)w(t)$ (since this case allows for the possibility of discounting, it may be considered fairly general). One might think that the result above does not hold because the TTO is time dependent while the SG is not. This is only partially true. The decision statistic values are:

$$\Delta\varphi^{SG}t = (av(q^A) + b)t^A - (av(q^B) + b)t^B = a\Delta\hat{U}(t^B/w(t^B)) + b(t^A - t^B), \quad (2.15)$$

which is exact only if $t^A = t^B$ in all applications or $w(t) = at$. In contrast,

$$\begin{aligned} \Delta\varphi^{TTO}t &= (w^{-1}(av(q^A)w(t^A)) - w^{-1}(av(q^B)w(t^B))) = \Delta\hat{U} \\ &\leftrightarrow w^{-1}(U) = aU + b. \end{aligned} \quad (2.16)$$

In both cases, exactness requires that w be linear. Concavity/convexity of w biases both measures upwards/downwards. The bias in the SG case depends only on the level of t , whereas in the TTO case it depends also on the level of q . Which generates the larger bias cannot be determined. The SG is only superior to the TTO if only morbid factors vary in a health state.

This result is quite startling since it suggests the SG, even with linear expected utility, cannot rank differences in health states correctly unless the conditions for the TTO to do so hold as well (note that it is not necessary to impose cardinal utility functions in the TTO case). This is because the TTO function is not a value function in this case (and cannot measure differences appropriately), and the SG does not value time appropriately relative to morbidity. The result is similar to that obtained in the

cost-benefit literature which shows the conditions for willingness-to-pay measures and money metrics to be exact are essentially the same: homotheticity of income in the indirect utility function. Instead of adding up across individuals, this decision statistic adds up across time periods, hence the difference in restriction.

QALY as an indicator of societal health Consider when QALYs are used to compare health profiles across large groups of people. The indicator used is $\sum_{i=1}^N \varphi^j(q_i) = \varphi^j(q)N$. This index has the same structure as the health status index above, except t is replaced by N , selfish preferences are replaced by social, and PE replaces TTO as the principle instrument.

Example: A commonly accepted and flexible expression for the social welfare function is $W = \frac{1}{N}(\sum_i v(q_i)^r)^{1/r}$. At $r = 1$ this is the inequality neutral SWF. Larger values for r indicate inequality affection, while lower values indicate inequality aversion. Then $\varphi^{PE}(q) = (\frac{v(q)}{v(\bar{q})})^r$, while $\varphi^j = \frac{v(q)}{v(\bar{q})}$ for all other instruments (assuming homotheticity in all other metrics). Then the bias may be expressed $B = (\varphi^j(\cdot) - \varphi^{PE}(\cdot))N = (\frac{v(q)}{v(\bar{q})} - (\frac{v(q)}{v(\bar{q})})^r)N$. Since $\frac{v(q)}{v(\bar{q})} \in [0, 1]$, $B = 0$ if $r = 1$, $B < 0$ as r falls and $B > 0$ as r rises (an unlikely situation). Thus, one would expect the bias to be non-positive.

Typically, under inequality averse preferences, non-PE instruments assign values to morbid states that are lower than those assigned by the PE instrument. In situations where the number of people occupying any given state may vary, the non-PE instruments overvalue states where there is inequality in outcomes and those with larger numbers of people. If these values are used to direct policy instead of PE values, too many resources will be directed to propagation and not enough to improving the health status of those who are already alive but unwell.

QALY's as a rationing device across patients Suppose there are only two types of prospective patients characterized by some high or low level of some factor, y (say, income). Suppose also that y is an enabling factor in the enjoyment of health (i.e. $U_{\xi,y} \neq 0$). Let this relationship be expressed $U(q, y) = v(q)w(y)$. Then $\varphi_H^{ES} = v(q)$, $\varphi_L^{ES} = v(q)(\frac{w(y_L)}{w(y_H)})$, and $\varphi_H^j = \varphi_L^j = v(q)$ for $j \neq ES$ (assuming homotheticity in the relevant metrics). Finally, for ease of demonstration, assume that social welfare is additive (this is a more restrictive function than the one above). The problem is to choose which patient gets the health improvement. The true value of a health status change is given by the SWF:

$$\Delta SWF(H) = (v(q^A) - v(q^B))w(y_H), \quad (2.17)$$

$$\Delta SWF(L) = (v(q^A) - v(q^B))w(y_L), \quad (2.18)$$

where $H(L)$ denotes that the high(low) individual experienced the change from q^B to q^A . Then one would choose the $H(L)$ individual if

$$(v(q^A) - v(q^B))(w(y_H) - w(y_L)) > (<)0. \quad (2.19)$$

By non-ES QALY analysis, one would obtain

$$\Delta \sum_1^2 \varphi^j(H) = v(q^A) - v(q^B) \quad (2.20)$$

$$\Delta \sum_1^2 \varphi^j(L) = v(q^A) - v(q^B), \quad (2.21)$$

indicating indifference as to who received the change. By ES QALY analysis, however,

$$\Delta \sum_1^2 \varphi^{ES}(H) = v(q^A) - v(q^B) \quad (2.22)$$

$$\Delta \sum_1^2 \varphi^{ES}(L) = (v(q^A) - v(q^B))(\frac{w(y_L)}{w(y_H)}). \quad (2.23)$$

Thus,

$$\Delta \sum_1^2 \varphi^{ES}(H) \geq \Delta \sum_1^2 \varphi^{ES}(L) \leftrightarrow \Delta SWF(H) \geq \Delta SWF(L). \quad (2.24)$$

Thus, only the ES instrument indicates the choice consistent with social welfare. Other instruments inappropriately discriminate against people who can derive greater benefit from a health status improvement.

The above examples all carry two lessons. First, the choice of decision statistic cannot be made without reference to preferences defined over the whole policy space. Arbitrary decision statistics cannot order states of the world appropriately, regardless of which QALY is used. Second, the solution to this problem of identification of the decision statistic may be circumvented if the QALY is redefined with a broader context (all relevant factors are included explicitly in the reference state) such that the QALY becomes a sufficient statistic over the whole policy space. This is basically the path Mehrez and Gafni (1989, 1991) have begun to follow for the first two cases above. The results are easily extended into other policy spaces.

2.2.7 Summary of Validity Results

The conclusions of this section are fourfold. First, necessary and sufficient conditions for validity of a QALY function exist and stem from the necessary welfare properties of the decision statistics which incorporate QALY values for program evaluation. Only necessary conditions exist for the QALY in those situations where the QALY is not a sufficient statistic (i.e. must be combined with other information to evaluate states). Second, any QALY index can be made exact if the state description in the survey is set equal to the actual policy situation in every characteristic. The conditions for exactness to be robust to changes in characteristic values differ across instruments, but while the conditions for some instruments are more stringent than for others, no

instrument is completely independent without some form of preference restriction. Third, because of this (variable) dependence on non-morbid factors, different QALY instruments do not, in general, generate the same QALY values. Fourth, the appropriate choice of QALY instrument depends on the choice of decision statistic in which its values are to be used, which in turn depends on the types of characteristics which change between the states under consideration.

Combining the last two points, one can conclude that the use of the wrong QALY function in a decision statistic is likely to lead to biased results and the wrong policy being implemented. The obvious conclusion to be drawn is that the researcher must choose his or her QALY instrument according to what is being evaluated. But this creates the dilemma of how to compare states which affect different characteristics when the evaluation of each state requires the use of different QALY functions which are noncomparable. Fortunately, the second result implies that exactness and comparability can be achieved in these cases by selecting one metric (i.e. one instrument) defined over an enhanced information set (i.e. expand the morbid state description to include not only the value of the metric, but the levels of all other factors which vary between states as well). In this way, correct decisions in health policy can be made.

2.3 Part II: Implementation

2.3.1 Nature of the Problem

The recent proliferation of medical procedures and increased demand for these services from an aging population have increased the need for health care programme evaluations.¹¹ Yet, despite the potential benefits to be had from such information,

¹¹Because of the peculiarities of the health care system, the price mechanism fails to allocate resources efficiently. Instead, efficient allocation must be undertaken by a central planner according

QALYs have been constructed for relatively few health states.¹² The premise of this section is that this discrepancy arises because, in most cases, the cost of a QALY analysis is too high to be recouped by the anticipated gains (over randomly allocated resources). This inference is supported by the fact that only high cost projects, where gains are apt to be large, have been evaluated to date.

There are several factors which contribute to these high costs. First and foremost, QALY values are not freely observable in the market, but must be obtained through expensive and time consuming surveys. Second, these surveys must be replicated across a large number of respondents because (a) responses to hypothetical survey questions tend to have high variances, (b) a representative sample of the population must be surveyed since the marginal consumer cannot be identified *ex ante*, and (c) a critical mass of QALY values is needed for meaningful comparisons of projects (there is no absolute critical cost per QALY value, so selection criteria can only be relative). Third, because health has many aspects, there exist many possible health states that have to be evaluated (see Williams [1983] or Boyle and Torrance [1984] for a discussion of how just a few characteristics and severity levels can generate a large number of distinct health states).

to the information provided by program evaluations. Torrance developed the QALY in the early 1970's as a health status index to be used for health care program evaluation. The premise of the QALY index is that time spent in poor health is worth less than time spent in good health. The QALY value is used to scale time alive according to the quality of the health endured during it, so that time spent in good health is assigned a higher value than time spent in ill-health. Since the index is based explicitly on preferences, it has a stronger welfare foundation than the other health status indexes available.

¹²QALY evaluations done to date have focused on cancer therapy (Sutherland et al. [1983] and McNeil et al. [1982]), major organ dysfunction (Weinstein and Stason [1977] and Churchill et al. [1984]), and neonatal care (Boyle and Torrance [1984]). Torrance (1987) and Maynard (1991) provide the most comprehensive QALY value tables, but even these are limited to about a dozen distinct states and tend to be drawn from the above three categories. Rosser and Kind (1978) have attempted to produce generalized QALY values based on morbid characteristics and not specific disease states. While this method may be an efficient solution, the use of only two vaguely defined characteristics is inadequate for many situations.

The usual response to this situation is that only high profile projects have been evaluated. Recently, some attention has been focused on how to produce QALY values more cheaply. The purpose of this paper is to examine the proposed cost saving techniques to determine what bias may ensue and its significance, and to find the most cost-effective method of obtaining QALY values to encourage their use. This is important because, if biased values are used, incorrect policy decisions (those that do not maximize welfare) may be made, and, if QALY based allocations are forsaken, allocation may become arbitrary (e.g. according to the lobbying effort of interest groups).

2.3.2 Literature Review

The choice of production technique is a two-step procedure. First, the instrument or method used to evaluate each state must be chosen. Second, the aggregation rule, which determines how many of these states need to be evaluated directly, must be selected. Both decisions involve trade-offs between the the quality of the QALY values produced and the costs of producing them. The literature review reveals these two parts of the production process have only been considered independently, even though the decision and outcomes are necessarily joint.

Instrument Selection

Preferences over health states are converted to a numerical scale by the use of a survey instrument, a device used to measure strength of preference between two states. While there are as many instruments as there are metrics against which to measure strength of preference, the analysis is restricted to a set of five which, between them, provide valid indexes for all current QALY applications (see the previous section). These are: category scaling (CS), standard gamble (SG), time trade-off (TTO), person

equivalents (PE) and extended sympathy time trade-off (ES).¹³ The different survey techniques/instruments require different amounts of time, effort and other inputs, and are associated with different cost structures. They also involve distinct decision processes, so they may yield different results. While quality-input trade-offs are obviously present, no cost function for instrument selection has been specified because the cost and output concepts are vague and are usually considered independently of one another.

The best sources for cost information are user manuals and surveys (such as Furlong et al. [1990], Drummond et al. [1987], and McDowell and Newell [1987]). Costs arise from the various inputs used in the evaluation process (see Drummond et al. for a description) and are difficult to calculate because the different inputs are often measured in dissimilar units and, to the extent common monetary values are available, these are often specific to the researcher's particular environment. Thus, most cost analyses resort to the use of imperfect and partial signals, such as time for respondent to complete task (Furlong et al.), response consistency (Froberg and Kane [1989]), and non-response rates (Patrick et al. [1973]). Such measures are often ordinally measurable, providing only a ranking of costliness (in descending order: SG, TTO, CS), but no quantification of these costs (although it is commonly believed the difference between SG and TTO is less than the difference between TTO and CS).

Apart from the lack of quantification, such analyses are inadequate because they do not consider the effects on the quality of the index produced associated with cost reduction, focusing only on the number of states valued and not how accurate these values are. Accuracy considerations can be found in an almost distinct body of literature dealing with the "validity" of various instruments. Theoretical concepts of validity are often ill-defined, resulting in partial and misleading analysis (e.g. Torrance

¹³The magnitude estimation and category scaling instruments are equivalent in this framework.

[1976a] and Loomes and McKenzie [1989] deal with only the interval properties of the function, focusing primarily on the SG; Pliskin et al. [1980] and Mehrez and Gafni [1989] examine context dependence, but only for the TTO).¹⁴

Such theoretical work has been accompanied by various empirical investigations into equivalence (see Stevens [1959], Torrance [1976b], Read et al. [1984], Wolfson et al. [1982], Bombardier et al. [1982] and Rosser and Kind [1978]). This body of work suggests that different instruments generate results that are highly correlated but not equivalent. While one must first identify the valid instrument before the significance of these differences can be ascertained, it can be concluded that some inaccuracy will arise when instruments are interchanged, suggesting a cost-quality trade-off exists.

Aggregation Methods

An aggregation rule is a procedure whereby values for states which have not been assessed directly (i.e. from a survey respondent) are obtained by some extrapolation from the values of states that have been assessed. The concept of valuation by select aspects was first discussed by van Praag (1968), but it was not until Keeney and Raiffa's book (1976) that due consideration was given to how these evaluations should be combined. Keeney and Raiffa establish the functional forms that must characterize utility in order for multi-attribute theory to be applicable¹⁵. These are the additive, the multiplicative, the multi-linear, and the holistic. Typically, each morbid characteristic is valued in isolation of any other morbid characteristic by direct means, and the value for any configuration of morbid characteristics is then reconstructed according to some function of the values associated with its component

¹⁴Conditions for validity are derived in the previous section

¹⁵While multi-attribute theory developed in economics, the theory of conjoint measurement evolved independently in the psychometric literature. Additive conjoint measurement is similar to additive forms of utility, while polynomial measurement resembles multiplicative forms.

parts. Different functions are associated with different aggregation rules with different parameter requirements (which entail additional valuations to identify).

Cost savings depend on the combination rule chosen. If K is the number of mutually exclusive characteristics, then the holistic approach requires 2^K QALY valuations, the multilinear approach requires $2^K - 1$, the multiplicative approach requires K and the additive approach requires $K - 1$.

Verification of the various independence conditions necessary for the different methods to produce identical values has been irregular. Such reconstruction methods are standard procedure for virtually all health status indexes (e.g. Activities of Daily Living, Sickness Impact Profile) and have been advocated by Torrance (1986) as a reasonable approximation to the holistic (direct) QALY method, yet very few authors have examined the underlying assumptions of such an approach and even fewer have attempted to verify these conditions empirically. Culyer (1976) recognizes explicitly the assumptions which make the atomistic (additive) approach viable in a QALY specific context. Rosser and Kind (1978) provide enough information for a crude test of additivity in their disability-distress matrix (the marginal disutility of distress increases with disability, refuting independence over these two characteristics), but do not perform such a test nor provide sufficient information to evaluate the statistical significance of the result. Giaque and Peebles (1976) employ multidimensional utility theory in the valuation of various states related to streptococcal sore throat and rheumatic fever (ten attributes), but only so far as to impose independence assumptions, not to test for them (with only 13 respondents, this is not feasible). Krischer (1976), in his study of cleft palate, finds that up to three-quarters of his sample of 119 exhibit pairwise independence of preferences over the three attributes of speech, appearance, and hearing. Although these results appear to contradict the evidence of Rosser and Kind, it must be remembered that the test procedures and the relevant

vector of characteristics differ between the two studies. Indeed, the biggest problem with studies of this nature is their specificity – the results are applicable to a very narrow set of morbid characteristics and are not useful for most applications.

Boyle et al. (1983, 1984) provide the first comprehensive and thorough analysis of atomistic methods in their work on neonatal intensive care. Faced with 23 characteristics that generated 960 possible states, they test whether the holistic values are additive functions of atomistic values. They find specifications with interactive terms are significantly superior to those without such terms. Thus, they adopt the multiplicative form of Keeney and Raiffa because it better accommodates these effects. Their analysis only tests for independence properties over extreme ranges of severity and, by the authors' own admission, only tests over a small subset of health states. They advise against the extrapolation of their results to other subspaces. A far more serious flaw in the analysis, one which they apparently do not recognize, is that the multiplicative form is never tested against more flexible alternatives: the alternative hypothesis of multiplicative separability is accepted even though the conditions necessary for it to be unbiased are never assessed.

Avenues for Further Research

This review demonstrates four gaps in the literature to date. First, the production processes are treated disjointly, even though this is infeasible in practice. Second, the objective function is not stated in terms consistent with the objectives of QALY analysis (welfare maximization). Third, the relationship between costs and the accuracy of QALY values is seldom made explicit. Fourth, meaningful estimates of the trade-offs involved are either unavailable or apply only to very specific contexts.

In this paper, trade-offs are modelled explicitly, with the bias generated by the choice of instrument appropriately nested within the bias function for the aggregation

method chosen. These bias functions are related to the welfare losses caused by inaccurate QALY values (this being a function of the expectation of choosing the wrong state and the consequences of this choice). Estimates of the trade-offs involved are provided for general health states that are useful in most applications.

2.3.3 Analysis

The purpose of this paper is to provide enough information about the cost-output trade-offs in QALY calculation so that the policy maker can make rational choices about which production methods to use. Because cost data are location specific, the production function, the primal to the cost function, is analyzed. The information provided is equivalent, but is robust to environments with different input prices.

Assumptions

The assumptions made in Part I regarding preferences and their domain are sufficient for the analysis here. However, a slightly modified notation for the decision statistics needs to be adopted to accommodate the additional QALY construction methods analyzed here. Hence, the assumptions of before are augmented by:

i) Let Ξ_i represent the “single-dimensioned” morbid state:

$\Xi_i \equiv (\bar{\xi}_1, \dots, \bar{\xi}_{i-1}, \xi_i, \bar{\xi}_{i+1}, \dots, \bar{\xi}_K)$ (i.e. Ξ_i is the vector of morbid characteristics with all but the i th element at perfect health).

j) The *holistic* QALY value for any health state q is obtained by surveying q with instrument j ($j = \text{CS, SG, TTO, PE, ES}$):

$$\varphi^j(q; t, \kappa).$$

(This corresponds to the QALY functions used in the previous section.)

- k) The *atomistic* QALY value is the (necessarily holistic) QALY value for a single-dimensioned morbid state:

$$\dot{\varphi}^j(\Xi_i; t, \kappa) = \varphi^j(\bar{\xi}_1, \dots, \bar{\xi}_{i-1}, \xi_i, \bar{\xi}_{i+1}, \dots, \bar{\xi}_K; t, \kappa).$$

- l) Let $\varphi^{j,A}(q; t, \kappa)$ denote the QALY function reconstructed by method A (A = multi-linear, multiplicative, additive) over atomistic functions derived with instrument j which is analogous to the holistic QALY value for the multi-dimensioned state q :

$$\varphi^{j,A}(q; t, \kappa) = f^A(\dot{\varphi}^j(\Xi_1; t, \kappa), \dots, \dot{\varphi}^j(\Xi_K; t, \kappa)),$$

where $f^A(\cdot)$ denotes the aggregation function for reconstruction method A .

2.3.4 Model

The policy maker seeks to obtain a set of QALY values which, when used in the appropriate decision statistic, correctly rank more preferred health care projects higher than less preferred projects. Such QALY values are called valid. QALY values which do not satisfy this property are biased.

QALY values are not freely available and the policy maker, faced with a limited research budget, may have to adopt cheaper and more biased methods of constructing QALY values than would be the case if QALY values were costless. The problem is to minimize the amount of bias incurred for a given cost saving, i.e.

$$\min_p B(p) \tag{2.25}$$

subject to

$$C \geq c(p) \tag{2.26}$$

where p is the production process, $B(p)$ is the aggregate bias of the QALY values generated by the production method p , $c(p)$ is the cost of producing these values by

production method p , and C is the evaluation budget.¹⁶ This problem is not actually solved in this paper, since the optimal choice depends on cost structures and the value of life, both of which are unknown.

2.3.5 Production Method

The production process consists of two parts: a survey instrument (j) and an aggregation rule (A):

$$p = \{j, A\}$$

There are five instruments available: the category scale (CS), the standard gamble (SG), the time trade-off (TTO), the extended sympathy time trade-off (ES), and the person equivalent (PE).

The CS requires that the survey respondent value a described health state relative to two reference states, say q^1 and q^0 . If such valuations are consistent with the utility function over certain outcomes, this instrument generates the following QALY function

$$\varphi^{CS}(q; t, \kappa) = \frac{U(q, t, \kappa) - U(q^0, t, \kappa)}{U(q^1, t, \kappa) - U(q^0, t, \kappa)} = aU(q, t, \kappa) + b, \quad (2.27)$$

where a and b are constants.

The SG requires that the survey respondent choose the probability vector for a reference or standard gamble involving given win and loss states ($q_{win} = q^1$ and $q_{loss} = q^0$ respectively) such that he or she is indifferent between accepting the gamble or the described health state with certainty. Then this instrument generates the following QALY function

$$\varphi^{SG}(q; t, \kappa) = p, \quad (2.28)$$

¹⁶The problem may be reformulated such that efficiency gains must recoup the costs of evaluation. This approach requires the monetary value of incorrect project selection be stated explicitly. In the case above, this value is implicit in the Lagrange multiplier.

where

$$\tilde{U}(U(q, t, \kappa), 1) = \tilde{U}(U(q^1, t, \kappa), p; U(q^0, t, \kappa), (1 - p)). \quad (2.29)$$

If the von Neumann-Morgenstern axioms hold

$$\varphi^{SG}(q; t, \kappa) = \frac{U(q, t, \kappa) - U(q^0, t, \kappa)}{U(q^1, t, \kappa) - U(q^0, t, \kappa)} = aU(q, t, \kappa) + b, \quad (2.30)$$

where a and b are constants.

The TTO requires that the survey respondent select how much time in perfect health (q^1 for convenience) is equivalent to a given period of time spent in the described health state. Then this instrument generates the following QALY function

$$\varphi^{TTO}(q; t, \kappa) = \frac{t - m}{t}, \quad (2.31)$$

where

$$U(q, t, \kappa) = U(q^1, t - m, \kappa). \quad (2.32)$$

If utility is homothetic in time, then

$$\varphi^{TTO}(q; t, \kappa) = \frac{\mu(q, \kappa)}{\mu(q^1, \kappa)} = a\mu(q, \kappa), \quad (2.33)$$

where a is some constant.

The ES is similar to the TTO, except the reference state is characterized by a given health state (q^1) and personal characteristics (κ_j). Then this instrument generates the following QALY function

$$\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) = \frac{t_j - m}{t_i}, \quad (2.34)$$

where

$$U(q, t_i, \kappa_i) = U(q^1, t_j - m, \kappa_j), \quad (2.35)$$

where i indicates the respondent's characteristics. If utility is homothetic in time, then

$$\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) = \frac{\mu(q, \kappa_i)}{\mu(q^1, \kappa_j)} = a\mu(q, \kappa_i), \quad (2.36)$$

where a is some constant.

The PE is based on social, rather than selfish, preferences, trading off the health status of groups of individuals to achieve states of social indifference. If social welfare, W , is welfarist, then this instrument generates the following QALY function

$$\varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) = \frac{N - m}{N}, \quad (2.37)$$

where

$$\begin{aligned} W(U(q, t_1, \kappa_1), \dots, U(q, t_N, \kappa_N)) = \\ W(U(q^1, t_1, \kappa_1), \dots, U(q^1, t_{N-m}, \kappa_{N-m}), \\ U(q^0, t_{N-m+1}, \kappa_{N-m+1}), \dots, U(q^0, t_N, \kappa_N)). \end{aligned} \quad (2.38)$$

If W is homothetic in population and $\kappa_i = \kappa$, $t_i = t \forall i$, then

$$\varphi^{PE}(q_1, \dots, q_N; t, \kappa) = \frac{U(q, t, \kappa) - U(q^0, t, \kappa)}{U(q^1, t, \kappa) - U(q^0, t, \kappa)} = aU(q, t, \kappa) + b, \quad (2.39)$$

where a and b are constants.

Except when preferences over morbidity are weakly separable from non-morbid factors and utility is homothetic with respect to all metrics, the different instruments can be expected to generate different QALY functions (see the previous section).

There are three aggregation methods available in addition to the holistic approach which is based on the entire configuration of the health state. These are described in Keeney and Raiffa (1976). Modifying their equations for the QALY case (where values are based on differences from a maximal rather than a minimal point), these are:

multi-linear If each characteristic is weak difference independent (WDI) of the remaining characteristics (i.e. the ordering of preference differences over characteristics does not depend on the level of characteristics which do not change), then $\varphi^j(q; t, \kappa)$ can be represented by a multi-linear combination of the atomistic QALY functions:

$$\begin{aligned} \varphi^{j,ML}(q; t, \kappa) = & 1 - \left(\sum_{k=1}^K \lambda_k (1 - \varphi^j(\Xi_k; t, \kappa)) \right) + \\ & \sum_{k=1}^K \sum_{h>k}^K \lambda_{k,h} (1 - \varphi^j(\Xi_k; t, \kappa)) (1 - \varphi^j(\Xi_h; t, \kappa)) \\ & + \dots + \lambda_{1,\dots,K} (1 - \varphi^j(\Xi_1; t, \kappa)) \dots (1 - \varphi^j(\Xi_K; t, \kappa)) \end{aligned} \quad (2.40)$$

with $\sum_k \lambda_k + \sum_{k=1}^K \sum_{h>k}^K \lambda_{h,j} + \dots + \lambda_{1,\dots,K} = 1$, where the λ 's are preference weights (or trade-off values) and K is the number of mutually exclusive morbid characteristics (assuming morbidity can only be present or not present for each characteristic (there are no severity levels within elements), this equals the number of elements in the set q).

multiplicative If each characteristic exhibits mutual preference independence (MPI) (i.e. preference for each characteristic is independent of the level of all other characteristics, or weak separability), then $\varphi^j(q; t, \kappa)$ can be represented by a multiplicative combination of the atomistic QALY functions:

$$\begin{aligned} \varphi^{j,M}(q; t, \kappa) = & 1 - \left(\sum_{k=1}^K \lambda_k (1 - \varphi^j(\Xi_k; t, \kappa)) \right) \\ & + \sum_{k=1}^K \sum_{h>k}^K \lambda \lambda_k \lambda_h (1 - \varphi^j(\Xi_k; t, \kappa)) (1 - \varphi^j(\Xi_h; t, \kappa)) \\ & + \dots + \lambda^{K-1} \lambda_1 \dots \lambda_K (1 - \varphi^j(\Xi_1; t, \kappa)) \dots (1 - \varphi^j(\Xi_K; t, \kappa)) \end{aligned} \quad (2.41)$$

with $1 + \lambda = \prod_{k=1}^K (1 + \lambda \lambda_k)$.

additive If each characteristic is strong difference independent (SDI) of the remaining characteristics (the actual value of the preference differences, not just the ordering, is invariant to the level of the characteristics that do not change, or strong additive separability), then $\varphi^j(q; t, \kappa)$ can be represented by an additive combination of the atomistic QALY functions:

$$\varphi^{j,D}(q) = 1 - \left(\sum_{k=1}^K \lambda_k (1 - \varphi^j(\Xi_k; t, \kappa)) \right) \quad (2.42)$$

with $\sum_{k=1}^K \lambda_k = 1$.

When the independence properties between the morbid characteristics do not hold, the different aggregation methods yield different values. The net difference between production methods depends on how the effects of non-independence between morbid characteristics interact with the effects of non-independence between morbid and non-morbid characteristics.

The budget constraint

Assume the policy maker's resources for evaluation are set at some level, C . The costs of producing a given set of QALY values by production method p is $c(p)$. This function is assumed to have the following structure

$$c(p) = L(A)c^j(q) \quad (2.43)$$

where $c^j(q)$ denotes the unit cost of evaluating any given health state, q , by instrument j ¹⁷ and $L(A)$ is the number of values that need to be found when aggregation method A is employed¹⁸.

¹⁷Furlong et al. (1990) report that respondents are able to evaluate five to six health states at a time. Thus, while there are significant fixed learning costs, these are quickly dissipated by respondent limitations. Efficiency dictates that the whole process be repeated frequently to achieve minimum costs. Because there are many states to be evaluated, replication will have to be undertaken and the average costs will appear to be relatively constant.

¹⁸All methods of reconstruction require the estimation of K atomistic functions (corresponding to the K mutually exclusive types and levels of health ailments), but different numbers of trade-off

The bias function

The bias function represents the losses the policy maker incurs when the estimated QALY values differ from the true or valid QALY values. In the previous section, utility theory was used to generate the conditions for a valid QALY index. It was found that different applications had different requirements which were fulfilled by different instruments. The difference between the values assigned to health states by the “true” index (denoted by $\varphi^t(q; t, \kappa)$) and the values associated with any other index are relevant only to the extent that they cause incorrect project rankings. This must be included in the specification of the maximand.

In the past, equivalence between the approximate index and the true index was assumed to be necessary for unbiasedness, regardless of how differences between the two indexes affected the decision statistic and, therefore, choices. The problem may be viewed from two contexts: *ex post* (true values are known), and *ex ante* (values are unknown).

In the *ex post* case, the QALY may or may not be a sufficient statistic (if it is, the QALY value alone determines the project choice; otherwise, the QALY value must be combined with other information before a choice can be made). If the QALY is a sufficient statistic, then the equivalence condition is overly strong since only the ordering of health states by the approximate and true indexes need be the same, i.e.

$$\varphi^{j,A}(q; t, \kappa) = \phi(\varphi^t(q; t, \kappa)), \quad (2.44)$$

where ϕ is an increasing monotonic function. In this case, the appropriate distortion

values (the higher order λ 's). Thus, the multi-linear method requires $(2^K - 1)$ parameters to be calculated, the multiplicative requires $(K + 1 - 1)$ such parameters, while the additive requires only $(K - 1)$. In comparison, the holistic approach requires 2^K valuations. Parsimony in these parameters is purchased with more stringent preference assumptions.

measure is ¹⁹

$$\phi^{-1}(\varphi^{j,A}(q; t, \kappa)) - \phi(\varphi^t(q; t, \kappa)). \quad (2.45)$$

If the QALY is not a sufficient statistic, equivalence is necessary (because of the measurability conditions imposed when QALY values are combined with other data), but is not sufficient since the welfare consequences of incorrect choices depend on the additional non-QALY information. The welfare loss is equal to this difference only if the QALY is cardinally related to utility (which may or may not be the case depending on the properties of the QALY and whether aspects of the state other than morbidity change). Since this information is specific to each assessment, the welfare consequences of a wrong decision cannot be evaluated until the nature of the assessment is identified and the sum of incorrect decisions must serve as a first order approximation of the losses generated by a wrong decision.

In the *ex ante* case, the policy maker does not know the true values of the health states under consideration, just the health outcomes themselves. This situation is more likely to arise than the *ex post* case – after all, if the true values are known, there would be no point in collecting the QALY data. The values for these health states may be viewed as random draws from the QALY interval (the set of all possible values). One may assume that the distribution of these values is uniform²⁰. Then, given that there are a large number of states to be evaluated, the probability that any one state is incorrectly ranked is a linear function of the difference between the

¹⁹ ϕ^{-1} may be found formally if the functional form of the two QALY generating functions are known, or by Box-Cox estimation if the only information available are the values generated by each function for some set of states.

²⁰The assumption of uniformity is not supported by the distribution of QALY values found to date (which appear to follow a Beta distribution). However, this may be an artifact caused by selection bias. Since there are not enough observations to confidently specify a distribution's parameters, and because the uniform distribution provides a good deal of intuition for the results obtained, and because the assumption probably holds over sub-intervals of the QALY range, the uniform density is adopted.

calculated and the true QALY values. (The larger is this difference, the larger the interval of neighbouring states against which this state is incorrectly ranked, the more likely the alternate health state will draw a value in this interval. Linearity arises from the assumption of random draws from a uniform distribution.) Then the probability of making a wrong choice for a single health state is

$$= a | \phi^{-1}(\varphi^{j,A}(q; t, \kappa)) - \varphi^t(q; t, \kappa) | + b \quad (2.46)$$

(ϕ^{-1} reflects the required degree of measurability), where $b = 0$ (since this probability must necessarily equal zero when the two functions are identical) and $a > 0$ ($a = 1$ only if the values of all states lie within the unit interval). The expected number of incorrect responses is the probability that any one response is wrong taken over all possible configurations of the ξ_k s, i.e. the sum of these probabilities.

If it is unknown what, if any, information is to be combined with the QALY values, then this sum represents an undominated expression for the welfare losses incurred when incorrect decisions are made (i.e. no other expression can be considered better). Thus, equivalence is undominated in situations of complete ignorance. However, if additional information regarding the true distribution of the QALY values in question, or how the values are to be subsequently used, becomes available, another representation of the welfare loss could be superior. This discussion indicates under what conditions equivalence is the appropriate criterion.

The policy maker's problem can be specified

$$\min_p \sum_{\{k\}} | \phi^{-1}(\varphi^{j,A}(q; t, \kappa)) - \varphi^t(q; t, \kappa) | \quad (2.47)$$

(where $\{k\}$ is the number of configurations of the morbid characteristics) subject to

$$C \geq L(A)c^j(q), \quad (2.48)$$

(the a above can be suppressed in the minimand if the Lagrangian is correspondingly adjusted).

2.3.6 Parameterization

In this section, parameter values for the above problem are derived for a broad range of health states. Such information is necessary to assess the cost-accuracy trade-off. While the policy maker is apt to know the nature of the constraint, having access to specific information about C and c^i , he or she is less likely to know the nature of the minimand. The main contribution of this paper is to provide a reasonable estimate of the minimand over a general range of health states so that rational choices can be made by the policy maker about how to obtain QALY values.

The bias generated by any particular production technique is a function of the biases arising from the two components of the production decision. Since the bias associated with the choice of instrument is nested within the bias associated with the aggregation rule, it is dealt with first in isolation and then in conjunction with the aggregation rule.

2.3.7 Bias from Instruments

Five QALY instruments are considered: category scaling, standard gamble, time trade-off, person equivalents, and extended sympathy time trade-off. The assumptions and functional representations of these instruments are discussed in the Analysis section.

From Part I, equivalence requires (1) independence of preferences over morbidity from non-morbid factors, and (2) homotheticity of preferences with respect to the metric variable (i.e. the variable against which the health state is measured must, for some monotonic transform, enter utility linearly). While empirical analysis is

necessary to verify these equivalence conditions, resources are not available to do so. Because these results are necessary to proceed with the rest of the analysis, they are reconstructed axiomatically and a sensitivity analysis is performed to assess how reasonable an approximation this approach provides.

Reference to the literature (see the review) suggests that independence is satisfied (comparisons show no disjointness in the relationships between instrument functions), but homotheticity is not (the functions are not perfectly congruent). Using this as a starting point, a utility function is posited with a functional structure and parameter values defined over a range consistent with empirical observation. While a great many assumptions must be made to support this function, it has been chosen so that none of them is very unreasonable given empirical evidence.

Consider the following utility functions which are able to accommodate discounting over time, non-linear expected utility (including non-independence in probabilities and regret theory), preference variation, and, in the case of social welfare, inequality aversion:

$$U(q, t, \kappa) = \int_0^t e^{-r\tau} \phi(p) q w(\kappa) d\tau \quad (2.49)$$

$$SWF = \left(\sum_{i=1}^N U_i^c \right)^{(1/c)} \quad (2.50)$$

(where the social welfare function is defined over all N members of society). In the simulation, the discount rate, r , is set to 0, 5, and 10 percent, a range which spans most reasonable estimates of time preference. Expected utility is expressed with a weighted probability function along the lines suggested by Chew (1980): $\phi(p) = \frac{p^a}{p^a + d(1-p)^a}$. This function can represent a variety of behaviours under uncertainty, including non-independence ($a \neq 1$) and regret theory ($d > 1$), as well as the standard von Neumann-Morgenstern case ($(a, d) = (1, 1)$). The range selected is $(a, d) = (1, 1), (.75, 1.5), (.5, 3)$, where $d = f(a)$ such that the responses follow the pattern

found in experimental settings by Kahneman and Tversky (1979). $w(\kappa) = \kappa$ is a generic function which scales the marginal utility of morbidity according to the level of the enabling factors, κ . The sub-utility function over morbidity is, for convenience, set equal to the level of q , which is assumed to be measurable as a single aggregate statistic (results generalize to richer specifications). The range of this function is restricted to $[q_{death}, \bar{q}]$ to avoid situations where the QALY functions could be undefined. Notice that this functional form assumes preferences over q are separable from non- q factors, a position that appears to be supported by the empirical evidence of Torrance (1976b), Read et al. (1984) and others that show mean orderings suffer no discontinuities, which would arise if this were not the case. The social welfare function is given by the mean of order c , with $c = 1, .5, 0$ (corresponding to inequality neutrality, and two cases with greater inequality aversion).

The QALY functions associated with these utility functions (after appropriate normalization) are:

$$\begin{aligned}
 \varphi^{CS}(q; t, \kappa) &= q \\
 \varphi^{SG}(q; t, \kappa) &= \frac{dq^{1/a}}{(1-q)^{1/a} + (dq)^{1/a}} \\
 \varphi^{TTO}(q; t, \kappa) &= -\frac{1}{rt} \ln(1 - q(1 - \exp\{rt\})) \\
 \varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) &= -\frac{1}{rt} \ln(1 - q(\frac{w(\kappa_i)}{w(\kappa_j)})(1 - \exp\{rt\})) \\
 \varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) &= q^c.
 \end{aligned} \tag{2.51}$$

The results of the sensitivity analysis are depicted graphically in Appendix B. The specification of ϕ affects only the SG function, the levels of r and t affect both TTO and ES, the level of κ affects the ES function, while the value of c affects PE. Divergence depends on which instruments are being compared and over what range of q . At times, differences can be quite large, even with conservative parameter values. While many conclusions can be drawn from this simulation exercise, probably the two most significant are that divergence tends to be greatest in the upper-middle range of

morbidity (where the most prevalent morbid states lie) and that differences between the instruments are evident even with conservative estimates of the parameter values. Inferences must be made with caution, however, since the simulation does not generate results consistent with some actual evidence. (The work of Sackett and Torrance [1978] suggests that ϕ has been misspecified and that $a > 1$. Since values at this level contradict the evidence of Kahneman and Tversky [1979] and the properties Chew [1980] writes the function ϕ should possess, the sensitivity analysis is not extended to this range.)

These results suggest that the biases from using inappropriate QALY functions can be quite significant, but are sensitive to the choice of parameter values.²¹ For the purposes of this paper, a given set of parameter values must be employed. These are chosen, where possible, to match empirical evidence in the literature, and, where a range of values is reported, small, conservative figures are chosen (this strengthens the final conclusions of the paper). These are: $r = .05$ (a value consistent with most reports of the social rate of discount (including Drummond et al. [1987])), $\phi(p) = \frac{p^{.75}}{p^{.75} + 1.5(1-p)^{.75}}$ (which mimics the pattern found by, among others, Kahneman and Tversky [1979]), $w(\kappa_i) = (1 \pm .25)w(\kappa_j)$ (i.e. preferences are allowed to vary around the “norm” by 25 per cent, an arbitrary but not unreasonable range), and $c = .9$ (to reflect a small amount of inequality aversion, again set arbitrarily but not unreasonably). These values seem to be a reasonable starting point for the rest of the analysis.

²¹It is straightforward to show equivalence is achieved if $r = 0$, $\phi(p) = p$, $w(\kappa_i) = w(\kappa_j)$ and $c = 1$ (i.e. when all metrics enter utility homothetically and independence with respect to the other variables holds). These values are not consistent with casual observation.

2.3.8 Bias from Aggregation

There are four methods of reconstructing QALY values: holistic, multi-linear, multiplicative, and additive. These are described, along with equivalence conditions, in the Literature Review (the holistic imposes no structure on preferences, the multi-linear requires weak difference independence (WDI), the multiplicative requires mutual preference independence (MPI), and the additive requires strong difference independence (SDI)).

A review of the literature shows what little work that has been done to verify these conditions is applicable to very specific health states, is often based on flawed test design (not all hypotheses are refutable), and considers only subsets of the strategy space (the aggregation choice is not nested within the instrument choice). Since previous work is unreliable or unsuitable, the independence conditions are evaluated in this paper.

Literature Review

The multi-attribute literature presents three methods of testing for preference restrictions directly (with utility rather than price and income data).

regression methods This method is based on the estimated relationship between the function values and the arguments of the function. The parameter estimates are checked to see if they satisfy certain restrictions (usually the analysis is confined to verifying the insignificance of higher order terms to support SDI, although other tests are theoretically possible). An example of this approach can be found in Klein et al. (1985).

Problems with this approach include: (1) the test is dependent on the model specification (and that the statistically best model is the true model), (2) that the

test may be sensitive to the scales of measurement of the independent variables (see Veit and Ware [1982] or Birnbaum [1973]), (3) the procedure tests for homogeneous separability, so that the test may reject the hypothesis of separability in separable structures because they are not also homogeneous (see Blackorby et al. [1978]), and (4) the parameter restrictions for the non-additive cases, particularly the multi-linear, are typically neither derived nor evaluated (especially in the health services research literature).

experimental methods This method requires the solicitation of values for both the atomistic and holistic value functions. The atomistic values are then aggregated, according to the weight restrictions, and compared to the holistic values. Boyle et al. (1983) provide an example of this approach. They set the λ_i 's (the preference weights used in the aggregation function) equal to the atomistic values and evaluate the expression $(1 + \lambda) = \Pi(1 + \lambda\lambda_i)$. They find λ is statistically different from zero, refuting the additive case, and, accepting this rejection of the one null hypothesis as support for the other hypothesis, adopt a multiplicative approach instead. Problems with their approach include: (i) the atomistic and holistic values are obtained by different methods (and, hence, may be on different scales), (ii) the atomistic values are evaluated only at the extrema (where distortions between instruments are most likely) so that values for severity are never calculated and (iii) only the additive case is posed as a refutable hypothesis. As with the regression approach, the tests are transformation dependent (so the caveat above in the regression section still applies).

axiomatic methods This method examines the underlying behaviour of preferences directly. Birnbaum's (1973) factorial methods are one example of this approach (although generalizations beyond the additive case are required). Closer examination

of the independence requirements reveals each independence condition generates a restriction on how utility may respond to changes in morbidity. For SDI, the size of the utility change caused by a change in some morbidity characteristic must be invariant to the level of any other characteristic, i.e.

$$\frac{\partial^2 U(q, t, \kappa)}{\partial \xi_i \partial \xi_j} = 0 \quad \forall i \neq j. \quad (2.52)$$

MPI requires weak separability across all morbid characteristics. This in turn requires that the marginal rate of substitution between any two characteristics be invariant to the level of any third characteristic, i.e.

$$\frac{(\partial U(q, t, \kappa) / \partial \xi_i) / (\partial U(q, t, \kappa) / \partial \xi_j)}{\partial \xi_k} = 0 \quad \forall k \neq i, j. \quad (2.53)$$

WDI requires that the ordering of utility differences be invariant to the level of any morbid characteristic not generating these differences, i.e.

$$\begin{aligned} U_i(\xi_i, \xi_j, \xi_{k \neq i, j}, t, \kappa) &\geq U_j(\xi_i, \xi_j, \xi_{k \neq i, j}, t, \kappa) \\ \leftrightarrow U_i(\xi_i, \xi_j, \xi'_{k \neq i, j}, t, \kappa) &\geq U_j(\xi_i, \xi_j, \xi'_{k \neq i, j}, t, \kappa), \\ U_i(\xi_i, \xi_j, \xi_{k \neq i, j}, t, \kappa) / U_j(\xi_i, \xi_j, \xi_{k \neq i, j}, t, \kappa) &\geq 1 \\ \leftrightarrow U_i(\xi_i, \xi_j, \xi'_{k \neq i, j}, t, \kappa) / U_j(\xi_i, \xi_j, \xi'_{k \neq i, j}, t, \kappa) &\geq 1, \\ \frac{\partial MRS_{\xi_i, \xi_j}}{\partial \xi_k} &\geq 1 - MRS_{\xi_i, \xi_j}, \end{aligned} \quad (2.54)$$

where the subscript on U denotes the first partial derivative with respect to that variable. Note that $SDI \rightarrow MPI \rightarrow WDI$. Note that SDI is transformation dependent.

All independence conditions are refutable. Complexity of the tests increases linearly with the number of characteristics, not exponentially (an additional characteristic requires additional tests be performed on that characteristic, but does not affect

the tests over the other characteristics). Tests can be done on subsets of characteristics without having to specify the relationship over all characteristics. However, test results cannot be interpolated between severity levels, so observations must exist for every configuration to be tested. Further, weaker versions of independence require more data points to be refuted.

Data

The data are comprehensively described in Appendix C. All variables are drawn from the General Social Survey (GSS) of 1985. This survey collected data on satisfaction with health and health status from a stratified sample of 11,000 Canadians.

Dependent Variable

The dependent variable chosen is satisfaction with health. This is the satisfaction concept most consistent with utility over morbidity, the variable of interest. Even so, additional assumptions are required. First, it must be assumed that this measure is cardinally equivalent to utility over morbidity or to utility over health after standardization for age (cardinality may be relaxed in the non-additive cases). This requires that responses be time independent (i.e. the satisfaction associated with a particular morbid state does not depend on the duration of that state) and that all individuals use the same time period (through which the morbid state endures) in assessing their well-being. It is also assumed that individuals assign the same meaning to the different categories (e.g. that “very satisfied” means the same thing to different people). Additional assumptions are required if structure in the preference relationships is used to infer structure in the QALY functions.

The above independence properties all apply directly to the function under consideration. In this project, however, verification of these independence conditions in

the QALY function must be done by reference to the properties of the utility function for which the data are available.²² Since $\varphi^j(q; t, \kappa) = f^j(U(q, t, \kappa))$ (see equation 2.51), what restrictions on $U(q, t, \kappa)$ are necessary for independence properties over q in $\varphi^j(q; t, \kappa)$ to hold (the necessity part of the following lemmata)?

Lemma 3: $\varphi^j(q; t, \kappa)$ exhibits WDI over q if and only if $U(q, t, \kappa)$ exhibits WDI over q .

Lemma 4: $\varphi^j(q; t, \kappa)$ exhibits MPI over q if and only if $U(q, t, \kappa)$ exhibits MPI over q .

Lemma 5: $\varphi^j(q; t, \kappa)$ exhibits SDI over q if and only if $U(q, t, \kappa)$ exhibits SDI over q and $\varphi^j \stackrel{c}{=} U$ (i.e. φ^j is cardinally equivalent to the utility function which is strongly additive).

Proofs: see Appendix A.

Because SDI is transformation dependent, there exists the potential for Type I error (i.e. there might exist some representation of preferences, $\phi(U)$, ϕ non-linear, which satisfies SDI when U does not). In this case, there exists some transform of the QALY data that can be aggregated and the ensuing sum can be untransformed to yield the appropriate QALY value. Depending on the form of this representation and how the QALY and U are related, these transformations can be quite complex.

In these cases, verification of reconstructability in U is sufficient to prove the possibility of reconstructability in φ . The functional form of the aggregator function is the same, although the parameter estimates can be expected to differ.

Independent Variables

The set of independent variables is determined by what is available in the data set and what is feasible for the statistical program used (SHAZAM). The GSS collects

²²In fact, satisfaction, not utility, data are available. It is assumed here that $S \stackrel{c}{=} U$ (see Sen [1985] for arguments why this relationship may or may not exist). The lemmata above are easily modified to be consistent with this relationship over satisfaction.

data on four categories of ill-health: chronic disability, including endurance (stair climb, walk, carry, stand), agility (bend, grasp, reach), and perception (hearing, sight); short term incapacitation (number of bed and sick days in the two week period prior to the interview), social health (number of contacts and visits with friends and family), and functional health (ability to carry out everyday activities). The last category is available only for persons over age 55, so it is dropped from the analysis. With a second order approximation, it is impossible to use all the variables in the analysis without exceeding the capacity of the statistical program. Instead, variables are grouped into five categories: endurance, agility, perception, short-term, and social (these groupings are suggestive of the conceptual frameworks often used to analyze health status so that this paper evaluates the appropriateness of such concepts in addition to its primary goal of assessing independence across morbid characteristics). If ill-health is indicated in any component of these aggregates, then the variable is assigned a value of one; otherwise, it is assigned a zero value. The significance of these groupings is tested on a ten per cent sample of the data by isolating one component and testing whether the coefficients on the isolated component and its respective interaction terms differ significantly from the coefficients associated with the aggregate variable based on the remaining components. Except for the agility variable, where the results become unreliable (coefficient values exceed the bounds required in a probability model), no statistically significant differences are found. Such an approach also helps to minimize the multicollinearity that is present in the data.

Specification

To verify the independence conditions, a hybrid of the regression and axiomatic approaches is employed. Basically, the utility function is estimated by regression techniques and the predicted values are examined to see if they are consistent with the utility restrictions derived from the axiomatic approach. The coefficients on the interaction terms are critical for these tests, so a flexible functional form is desirable. To curtail the profusion of parameters, however, only a second order approximation is used (this assumes higher order terms are insignificant). Since the multilinear expression is essentially a second order approximation, it can not be refuted in this specification. The tests are evaluated anyways, more as a demonstration of methodology than for the results themselves.

Since the independent variables are characterized by binary structures²³, the best choice of flexible functional is the Generalized Leontief (GL).^{24 25 26} Thus, the true relationship to be estimated is

$$S = \beta_0 + \sum_{i=1}^5 \beta_i q_i^{0.5} + \sum_{i=1}^5 \sum_{j>i}^5 \beta_{ij} q_i^{0.5} q_j^{0.5} + R \quad (2.55)$$

(where the third and higher order terms, included in the remainder R , are assumed

²³Note that this ensures Birnbaum's (1973) criticisms are unjustified (because there is no measurement scale that can distort the model specification).

²⁴Of the alternative specifications, the translog has undefined terms because of variables with zero values (and if these values are rescaled away from zero, the estimates will be biased since the translog is sensitive to measurement scale); the generalized CES requires evaluation of differences from base levels, but levels are a meaningless concept with ordinal data.

²⁵Actually, the binary structure of the independent variables poses problems for any flexible functional form since derivatives are not well-defined. As such, a Taylor's series expansion, on which most flexible functional forms are based, is not valid for these data. However, if Diewert's (1973) original definition of flexibility is reinterpreted in a discrete framework, some support is given to the flexibility of the specification.

²⁶Even without the structural problems of binary data, the GL is not truly flexible since it imposes homogeneity (see Blackorby et al. [1978]). It is not possible to rectify this problem with the data available since (1) higher order approximations exceed the capacity of the statistical package used and (2) the data are not generated by any optimizing behaviour so the function cannot be partitioned into variable and scale components.

to be insignificant and not to affect the estimates of the lower order terms).

Estimation

Since the dependent variable is an ordered categorical measure, taking on one of four ranked values, the method used to estimate the utility relationship is an ordered probit analysis.²⁷ The ordered probit assumes normality in the distribution of errors. While the Central Limit Theorem does not hold in this case (because the individuals choose the “state” they are in, the independence conditions are not satisfied), the ordered probit is still chosen over the alternative distributions (e.g. the ordered logit) because it is not clear that the data are better represented by any alternate extreme value distribution and because the probit is easily modified to represent ordered rather than unordered data. It is also assumed that the errors are homoskedastic. These two very strong assumptions about the distribution of errors are necessary if the ordered probit estimates are to be used to represent preferences.

Censored analysis is not necessary since response rates for both the dependent and independent variables are high and the collection of data is well stratified to represent all segments of society. Curvature need not be imposed since the function to be estimated is not the result of optimizing behaviour.

Instead of estimating the true relationship directly, ordered Probit analysis estimates the probability that a response will fall into a particular category if the responses are generated by this relationship. Denote the true relationship as $S(q) = f(b^T q)$. If the i th observation falls into the j th category, assign z_{ij} a value of one;

²⁷Ordinary regression methods impose the condition that differences between levels are equal. Unordered probit methods fail to recognize that responses exist on the same continuum. Ordered probit methods not only recognize both facts, but can be used to estimate the differences between the levels so that the ordinal information provided by the dependent variable can be converted to numerical values.

otherwise, assign it a value of zero. Then the log-density for a single observation is

$$P(z_{ij} = j) = \sum_{j=1}^J z_{ij} \log(\Phi(\alpha_j - f(b^T q)) - \Phi(\alpha_{j-1} - f(b^T q))), \quad (2.56)$$

where Φ is the normal cumulative distribution function on which Probit analysis is based, and α_j is the estimated step function measuring the difference between the j th and $(j - 1)$ th categories (although there are four categories, there are only two step functions because the density of the last category is one less the sum of the other three). The log-densities summed across all observations generate the likelihood function which is estimated by maximum likelihood methods. The functional form of $f(b^T q)$ is the second order Generalized Leontief (equation (2.55) above).

Estimates

Estimation is done with the SHAZAM statistical package as a non-linear regression problem using numeric derivatives and the Davidson-Fletcher-Powell algorithm. While the coefficient estimates are very robust to the choice of starting values, the standard errors are sensitive to the number of iterations needed for convergence (this is typical of the algorithm employed to calculate the information matrix). Since the basic starting values employed (based on linear approximations) do not lead to convergence until after twenty or more iterations, one may assume the approximations of the covariances are reasonable.

Of greater concern is the sensitivity of the results to the inclusion of certain observations. Estimates are sensitive to the inclusion of respondent groups which are characterized by particular morbidity profiles (e.g. the elderly who are typically in poorer health than the general population, farmers who are in good health or select out of the occupation). This could indicate preference variation or, more likely, under-represented cells. For this reason, the entire data set is used for the final estimates.

The results for the entire sample are presented in Appendix D. Examination of the data suggest satisfaction with health may contain a longevity component as well as satisfaction with morbidity and that these variations fall under three groupings: the young, the middle-aged, and the elderly.²⁸ The demarcation for the young is at thirty years, while the cut-off for the elderly can be put anywhere between 55 and 65 years (the pattern of preference variation is discernible at the lower value, but does not achieve significance until the higher value; the upper cut-off is chosen.) The estimation is performed on these sub-groupings as well (see Appendix D for results).

Probit estimation yields the results presented in Appendix D. Since the satisfaction variable takes on higher values the more dissatisfied the respondent is with his or her health, and the health variables take on positive values in the case of ill-health, one would expect the first order coefficients (E, A, P, S, and L) to be positive (i.e. that ill-health would be associated with higher levels of dissatisfaction). Over the entire data set, all first order coefficients have the expected sign, and all but social ill-health are significantly different from zero (in all tests, the critical significance level is set at 5 per cent).

Among the higher order terms, one would expect positive signs if the health components are substitutes (e.g. loss of one faculty makes a person more reliant on other abilities) and negative if they are complements (e.g. if one needs a combination of abilities to perform a certain function, then the marginal disutility of losing one ability is less if the other ability is absent than if it is present). Six of the interaction

²⁸Preference variation by socio-economic characteristics is found when a Hausman-Wise (1978) residual test on the original specification is performed. The most significant variables are age and occupation, followed by ethnicity, marital status, and household size. The age effect suggests the satisfaction with health variable is not perfectly congruent with satisfaction with morbidity and that age standardization is called for. The other effects, with the possible exception of ethnicity, can be explained in a simple human capital-time use model. While conclusions can be drawn more confidently if standardization is performed on these characteristics as well, this is not feasible because further subgroupings result in very poor representation in some cells (e.g. there are few young farmers in poor health) and this generates unreliable results.

terms are positive in sign (half being significant), while four are negative (none being significant). The results suggest, among other things, that individuals who lack communication faculties (P) compensate by relying on social contacts (L) and vice versa. The set of interaction terms is significant overall (Wald test of 43.2 with ten degrees of freedom).

Looking at the estimates in Table D.1 of Appendix D, it is clear that they vary across the three age subgroupings.²⁹ In particular, the estimates suggest substantial variation exists over the parameter estimates for short term and social ill-health, and that the elderly are, *ceteris paribus*, more satisfied with their health (because the constant term is smaller). Given that the estimates are robust to choice of starting values, these differences can only be explained as caused by some underlying difference in the preference ordering (e.g. conditioning on longevity) or by biased estimates resulting from poorly represented cells (e.g. the young tend to be in very good health, the elderly are often in poor health). Coefficient patterns tend to replicate across the three sub-groups, but with lower statistical significance (probably due to the smaller sample sizes), suggesting the latter reasoning may be more correct.

Tests

The procedure is to test the parameter estimates to see if they significantly differ from the following restrictions which are generated by the respective independence conditions associated with each aggregation rule:

²⁹LR tests indicate the sub-equation estimates for the young and elderly differ significantly from those for the equation estimated over the whole data set. The likelihood ratio test is constructed by comparing the log likelihood for any specific sub-group of data with the parameter estimates from the whole data set imposed, against the log likelihood for that sub-group of data evaluated at the parameter values found when only the sub-group data are used in the estimation. Test values for the young and the elderly are 84.99 and 138.33 respectively (distributed Chi-squared with 18 degrees of freedom), while the middle-aged equation does not differ significantly (test value of 22.57).

- i) $\partial^2 U(q, t, \kappa) / \partial \xi_i \partial \xi_j = 0 \leftrightarrow \beta_{ij} = 0 \forall j \neq i$ ³⁰ for SDI (10 conditions).
- ii) $\partial MRS_{\xi_i, \xi_j} / \partial \xi_k = 0 \leftrightarrow \beta_{ij} - (\frac{\beta_j}{\beta_k}) \beta_{ik} = 0 \forall k \neq i \neq j$ for MPI³¹ (45 conditions) ³².
- iii) if $\partial U(q, t, \kappa) / \partial \xi_i \geq \partial U(q, t, \kappa) / \partial \xi_j$ ($\beta_i \geq \beta_j$), then
- $$\frac{\partial((\partial U(q, t, \kappa) / \partial \xi_i) / (\partial U(q, t, \kappa) / \partial \xi_j))}{\partial \xi_k} \geq 1 - (\partial U(q, t, \kappa) / \partial \xi_i) / (\partial U(q, t, \kappa) / \partial \xi_j)$$
- $$\leftrightarrow \beta_{ij} - (\frac{\beta_j}{\beta_k}) \beta_{ik} \geq \beta_k - \beta_j \forall k \neq i \neq j \text{ for WDI}$$
- (30 conditions).

The parameter restrictions for SDI are tested using t-tests on the coefficients on the individual interaction terms (the β_{ij} 's), and Wald statistics are used to test the joint restrictions. The other conditions (for MPI and WDI) require non-linear tests (since they involve combinations of the coefficients) and a Wald test is chosen for these cases. (The advantage of the Wald test is that it requires only unrestricted estimates (saving the non-trivial resources needed to re-estimate the equations with restrictions imposed). The disadvantage is that the test is dependent on the representation of restrictions (although the structure of the problem suggests the representation used is the most reasonable). Because the number of observations is large, asymptotic equivalence between this test and the LR and LM tests may be assumed (when the LR test is used to evaluate the SDI conditions, the resultant test statistic differs from the

³⁰The estimated coefficients are not the true marginal utilities, but rather the incremental probability effects. The marginal utilities may be recovered by premultiplying these values by the appropriate probability distribution weights. However, these weights are the same within each test restriction and cancel out. Therefore the tests may be performed on the estimated coefficients without adjustment.

³¹The point of approximation is assumed to be perfect health.

³²The MRS relationship described in the text generates 30 conditions, while the approach actually used (a pairwise test of the interaction terms to determine if the scale factors, λ , necessary for $\beta_{ij} = \lambda \beta_i \beta_j$ to hold, are statistically equivalent across all pairs of interaction terms) generates 45. The conditions are redundant in one direction: if all 30 of the MRS conditions hold, then all 45 of these hold as well; if none of the 30 hold, some of the 45 can still hold, in which case some cost saving combinations can be made. Theoretically, this case is described by $U(\xi_i, \xi_j, \xi_k) \neq U(v(\xi_i, \xi_j), \xi_k)$ and $U(\xi_k, \xi_l, \xi_j) \neq U(v(\xi_k, \xi_l), \xi_j)$, and $U(\xi_i, \xi_j, \xi_k, \xi_l) = U(v(\xi_i, \xi_j), w(\xi_k, \xi_l))$.

Wald's in only the second decimal place and does not affect the conclusions drawn). Thus, the results of the Wald tests may be accepted with confidence.) In addition to these individual cases, it is also interesting to know if the restrictions hold jointly, both overall and with respect to some important subgroupings (such as whether the components of chronic ill-health are independent of one another³³, whether the three main groupings of ill-health are separable from each other³⁴) which indicate when certain types of ill-health can be evaluated in isolation of the other categories. This requires that the independence conditions hold over all morbid characteristics within the relevant group (e.g. to test whether short term ill-health is separable from chronic ill-health, one would have to test jointly the hypotheses that short term ill-health is separable from endurance, that short term ill-health is separable from agility, and that short term ill-health is separable from perception, since endurance, agility, and perception describe chronic ill-health in this model). Such joint tests are also more reliable than the individual tests because they are based on more information.

Because parameter estimates vary according to age, independence tests are performed for the whole sample (this is the more reliable equation if the differences arise from representation problems), and for each of the three subgroups (if the differences are due to preference variation, then these equations yield the more valid results).

Additive Tests

SDI is evaluated by examining the second order coefficients (SDI is refuted if they differ significantly from zero).³⁵ There are ten conditions to evaluate. Table D.2 of

³³This assumption is commonly made in the construction of other health status indexes which measure the value of a health state by summing the component values.

³⁴There is a common demarcation of health according to physical, social, and emotional well-being (see Torrance [1986] for a discussion); it is also common to treat chronic ill-health in isolation of periodic ill-health episodes, although the justification for this has never been clear.

³⁵Actually, reference must also be made to step values between categories since these indicate what transform of the response variable obeys SDI if the higher order terms are insignificant. Assuming the higher order terms are insignificant, then, if the steps are equally spaced, AI is supported, but

Appendix D reveals 30 per cent of these restrictions are violated for the entire sample. The t-test associated with the variable ExA (the cross term on endurance and agility) is greater than the critical value (at 2.48), suggesting that the value of (E+A) will misrepresent the true value of the state where both E and A are present (since the coefficient is positive, the predicted level of satisfaction is too high). Conversely, the t-test for the variable ExL is .526, which indicates the sum of the individual values for E and L does not differ significantly from the value of the state where both E and L are present.

Across categories, it is clear that the additive structures are not supported, with the chronic and short variables having the strongest result (a Wald test statistic of 21.38, see Table D.3 of Appendix D). This result also drives the significance of the more inclusive tests (all characteristics and the tri-category test). The only joint tests which are insignificant are those which separate only social ill-health from a given category. This is not surprising since the social ill-health variable itself is insignificant, the high standard errors allowing a broad range of acceptable proxy coefficients. Within categories (i.e. within chronic ill-health), the additive structure is marginally rejected.

Results suggest that simple additive structures across the categories studied here impose significant bias on the QALY values obtained, and are doubtful within the chronic disability category (although the results hold only for affine transformations of the satisfaction measure). Since the cost savings associated with the additive case can only be achieved by accepting significant bias, the other aggregation methods should be assessed to see if unbiasedness can be attained with more moderate cost savings.

if not, AI holds only for some non-linear transform of the response scale.

Multiplicative Tests

MPI is evaluated by examining whether the scaling factors on the first order coefficients which produce the higher order coefficients differ from each other (e.g. if $\beta_{ExA} = \lambda^1 \beta_E \beta_A$ and $\beta_{ExP} = \lambda^2 \beta_E \beta_P$, one asks if λ^1 and λ^2 differ significantly from one another). There are 45 such combinations to evaluate. Table D.4 of Appendix D reveals 24 per cent of these restrictions are violated. For instance, the Wald test associated with the pairs (ExA) and (ExP) is 4.62, which is significant when compared to the appropriate chi-squared value. Thus, the use of the scaling factor derived from (ExA) on E and P will produce a biased value of (ExP). Conversely, the test value for the pair (ExA) and (ExS) is 0.001, which is clearly insignificant, so that the use of the scaling factor derived from (ExA) may be used on E and S without biasing the value of (ExS). Note that the PxL variable (perception and social ill-health) is a focal point for rejection.

Individual test results seem to suggest that the multiplicative structure is not much better than the additive. However, in the joint tests, multiplicative structures cannot be refuted (the joint test over all categories could not be performed because the number of restrictions exceeded the number of parameters – however, all the sub-tests, which together encompass all restrictions contained in the global test, do not reject and, except in extreme cases (see Kennedy [1985]), this result implies the global restriction will also be satisfied). As a demonstration, consider the tri-category test, which cannot be performed. The components of this test are chronic-short, chronic-social, and short-social. The first two are given in Table D.4, the last in Table D.2 (SxL; the multiplicative test is not defined over two variables, but is a weaker version of the additive test). None of the three sub-tests is significant, so it would be surprising, though not impossible, for the tri-test itself to be significant. Furthermore, PxL is a focal point of rejection in the individual tests, but since these are characterized by

very high standard errors, the joint tests should not weigh these rejection points as heavily as the more definite non-refutable points. The joint tests should then be more inclined to the non-refutable cases. Since joint tests are based on more information than the individual tests, assigning more weight to the less variable restrictions, they are usually assumed to be more reliable. The multiplicative test within the chronic ill-health category could not be performed because of singularity. This singularity simply reflects the estimates nearly satisfy additive independence. Because the conditions for additive structures are more stringent than for multiplicative structures, it can be said that mutual preference independence cannot be rejected because the conditions for strong difference independence cannot be rejected.

These results suggest that multiplicative structures may be used across the categories studied here without significantly biasing the QALY values obtained. Since the multiplicative structures are nearly exact, the significant extra costs of the less restrictive aggregation methods cannot be easily justified. Also, the sub-aggregates of physical and social health, and short and long term health that are frequently used in the literature are proper aggregate commodities (see Deaton and Muellbauer [1980]) and may be treated as distinct aspects of health.

Multi-linear Tests

WDI is evaluated by examining a different combination of coefficients than MPI. It is evaluated by Wald tests. Since the second order approximation used to estimate the coefficients imposes the multilinear structure on the satisfaction relationship, rejection of any of the restrictions is theoretically impossible. Table D.12 of Appendix D reveals any combination which is significantly different from zero has a sign that is consistent with the restriction (unlike SDI or MPI, these are one-tailed tests).

Caveats and Implications

The test results suggest that health does indeed consist of aggregate commodities (lending partial support to the World Health Organization's conceptual framework) and that these may be used to reduce substantially the costs of valuing multidimensional health states. Results also suggest that simple additive structures are inappropriate across the categories studied here and are doubtful within the chronic disability category (although the result holds only for affine functions of this satisfaction measure). This result suggests indexes like the ADL (activities of daily living), which sum the values of various health characteristics to obtain a value for a health state, are probably biased. Certainly, a multiplicative format would yield more reliable results, yet it appears not to have been considered despite the little extra effort involved.³⁶

Some cautionary notes do need to be made, however. First, these results are based on average preferences. There is some evidence that the representative consumer is a fallacious concept in this situation. More experimental work, like that of Krischer (1976), needs to be done to see if the results hold at the individual level. Second, while these results hold over general variables over which some interpolation is possible (i.e. individual characteristics within distinct groups should still obey the patterns exhibited by their respective aggregates), there are still relationships within these aggregates that need to be specified. Furthermore, emotional health is not included in the analysis, so this can be considered only a partial analysis of the usual demarcation of health into physical, emotional, and social well-being. Third, the test results for WDI are vacuous since only a second order approximation is used. Higher order

³⁶In the sub-groups, the rate of violation is lower, although the pattern of violations is preserved in the two older groups. It is assumed that this reflects sample size problems, and that the underlying preference relationships are the same as for the whole sample.

approximations, even if possible, are not called for since, because MPI cannot be rejected, one can assert that the weaker condition of WDI also cannot be rejected. Fourth, the significance of the results may have low power in both an empirical and a theoretical sense.³⁷ The Blackorby et al. (1978) criticism that the test might be overly strong because it depends not only on independence, but homogeneity as well, is not of great concern for those hypotheses that are not refutable. Only SDI is rejected and, as is shown below, the QALY values generated by an additive aggregation rule are all biased regardless of the transformation (instrument) employed (the QALY instruments restrict the set of allowable transforms as well, further weakening the criticism).

2.3.9 Synthesis

The purpose of this section is to synthesize the results obtained so far by integrating the bias functions with each other and with costs. While the Modelling section of this chapter describes how bias is generated by the various production methods, it can not indicate how large the differences generated are. This section also deals with the fourth caveat above by demonstrating that the amount of bias caused by additive

³⁷The grouping of variables artificially linearizes the utility function which is then estimated in linear segments. However, the number of independent variables is large enough to make such segment-wise approximations fairly close to the true function. The tests are carried out only at the point of approximation. In theory, this point is assumed to be perfect health, although empirically this assumption is false. In either case, acceptance of the independence condition has low power since the test result only indicates that independence holds at this one point, but not necessarily any other. Rejection of the condition is robust, however, since, if independence does not hold at some point in the function, it obviously cannot hold over the whole function. Statistically, the tests have low power (none of the individual rejected hypotheses had power exceeding 40 per cent in the interval around the calculated value). This result may be due to mis-specification (the data do not fit the normal distribution well) or to the categorical nature of the dependent variable (the imprecision of the response creates problems typical of small samples). For a majority of the individual restrictions (60 per cent), acceptance of the null is supported more than rejection (when the probabilities are set equal to each other, the calculated values lie outside the rejection region). One would expect the joint tests to have higher power since they are based on more information.

structures does not vary significantly across the transformations associated with the different QALY instruments.

A parameterization of the bias function is generated using simulated QALY values for the fifteen health states in the model specified in the Empirical section (see Appendix D). The QALY values are constructed according to the specifications given in the Modelling section. The expected satisfaction estimates from the Empirical section of this paper are transformed so that the values generated are positively related to satisfaction. The transformation used is

$$v(q) = \frac{\hat{S}(q) - \hat{S}(\underline{q})}{\hat{S}(\bar{q}) - \hat{S}(\underline{q})}, \quad (2.57)$$

where \hat{S} is the predicted satisfaction response, $v(q)$ is the transformed predicted response, and \bar{q} and \underline{q} are the perfect health (no ill health in any category) and worst health (ill health in every component) states respectively. Assuming $T = 20$ and $v(q) = \mu(q, \kappa)$, the utility function over certain outcomes is:

$$U(q, t, \kappa) = \int_{t=0}^{20} \exp\{-rt\} v(q) dt. \quad (2.58)$$

This expression is substituted into the instrument functions to generate $\varphi^j(q)$, or the holistic QALY values, using the arbitrarily set parameter values described in the Empirical section under instrument generated bias ($r=.05$, $\phi(p) = \frac{p^{.75}}{p^{.75} + 1.75(1-p)^{.75}}$, $.75\mu(q, \kappa_i) = \mu(q, \kappa_j)$, and $c = .9$) and common reference points of perfect health (q^1) and worst health (q^0).³⁸ These are given in Table D.14 of Appendix D.

The QALY values associated with single morbid characteristics are then used to generate values under the different aggregation regimes (the trade-off values may be

³⁸This level is implicitly assumed to be a death equivalent (i.e. $v(\underline{q}) = 0$). If it is not, then the simulated QALY values will not be mapped to the usual interval and will not be comparable to QALY values reported in the literature. The results in this paper will be internally consistent regardless.

found from the values for single aspect states because of parameter restrictions).³⁹ The multilinear structure is identical to the holistic structure in this specification. The calculated values are given in Tables D.15 through D.19 of Appendix D.

The results should be accepted cautiously since the significance of the differences is not accounted for (significance cannot be incorporated since the instrument values are not generated empirically and therefore are not associated with any margin of uncertainty)⁴⁰. The results show that the ordering of states is preserved across all instruments, but not across the different aggregation methods ((ExP,ExL) and (AxP,AxL,PxL,SxL) are reversed in the CS and PE estimates, (ExP,ExL,AxS) and (AxP,AxL,SxL) are reversed in the SG estimates, and (ExP,ExL,PxS) and (AxP,AxL,PxL,SxL) are reversed in the ES and TTO estimates). This result stems from the fact that the utility function imposes separability of morbid characteristics from non-morbid characteristics (so there exists an appropriate transform, ϕ^{-1} , that generates equivalence). When equivalence is required, the appropriate measure of bias is the sum of the absolute deviations of the estimated values from the values of the “gold standard” or true production method over all health states defined for every instrument.⁴¹ Results suggest that substitution across aggregation methods (read down the columns in Tables D.20 through D.24 of Appendix D) typically involves less bias than substitution across instruments (as read across the rows in the same

³⁹The trade-off value for the multiplicative structure is found by solving the polynomial restriction associated with such structures (see the Modelling section). Only one non-zero real root within the range allowed exists for each of the instruments. Except for the SG, the values are in line with those found by Boyle et al. (1983).

⁴⁰The biases resulting from multiplicative aggregation are insignificant, while most of those from the additive reconstruction are probably significantly different from zero.

⁴¹Recall that the choice of “gold standard” depends on how the QALY is used to evaluate projects and that different QALY functions are appropriate in different contexts. Thus, differences are taken with respect to each instrument’s holistic values so that the bias of using any one production method in any particular context can be assessed.

tables). Note that the bias effects from instrument substitution and aggregation technique may offset each other. These results are due to a combination of the estimated preference structure between morbid characteristics and the imposed curvature on the non-morbid characteristics.

Additive structures do generate less bias for those instruments which are non-linear transforms of the satisfaction response, but in no case does this bias fall by more than ten per cent. This suggests that the dependence of the test for SDI on homogeneity does not decrease its validity for the QALY applications considered here.

The choice of how much bias to accept should not be made without consideration of cost savings. Since very little bias arises from the use of multiplicative aggregation structures and the cost savings are substantial (in this case, $c^j(q)10$, where $c^j(q)$ is the unit cost of evaluating any health state with instrument j), it is hard to justify the holistic approach. If only orderings of morbid states are required, then the cheapest instrument should also be adopted (resulting in savings of up to $(\max_j\{c^j\} - \min_j\{c^j\})5$ and no additional bias). If equivalence is required, as is more likely to be the case, the bias-cost trade-offs have to be considered: switching instruments saves $(c_{max} - c_{min})5$, but results in additional bias as read across rows in the simulation table; adopting an additive aggregation structure over a multiplicative structure saves c^j1 , but results in additional bias as read down the columns in the same table. In general, the more health characteristics that are valued, the greater the absolute cost savings of switching from multi-linear to multiplicative aggregation structures, and the greater the relative cost savings of switching instruments rather than aggregation structures from multiplicative to additive. The decision about whether and how much bias to accept ultimately depends on the expected costs from making the wrong project choice (this information is summarized in the lagrange multiplier) and how this measure of bias increases the likelihood of making such a choice (the relationship is linear only

in large samples).

One of the problems with QALY analysis is obtaining the QALY values themselves. This process is costly, so much so that QALY-based evaluations can seldom be justified. If the policy maker is willing to accept reasonable approximations of these values (which might be sensible given the high standard errors associated with empirical QALY values), cost saving strategies are available.

This chapter assesses these strategies. It improves on earlier work by (1) fully specifying the strategy space, (2) identifying the loss functions in terms consistent with the goals of QALY analysis, (3) testing for the least well known separability conditions over broadly defined morbid states, and (4) partially quantifying the trade-offs involved in the different strategies.

Conclusions include that preference structures support the use of multiplicative, but not additive, aggregation techniques (a by-product of this result is the justification of commonly held but unsubstantiated conceptual frameworks of health based on grouped characteristics). The result is of significant practical importance since it suggests substantial cost savings in QALY construction can be achieved without generating distortions in the QALY index, and that other health status indexes based on additive structures can be substantially improved at slight additional cost by adopting a multiplicative aggregation rule. This result suggests the most commonly invoked cost-saving strategy (instrument substitution) is sub-optimal. Unlike all previous work in this area, which focuses on and applies to only specific health states, these results hold over broadly defined and nearly comprehensive health categories. Results may therefore be generalized to specific characteristics which lie within these categories.

Additional cost savings can only be achieved with some amount of bias. The researcher's decision at this stage will depend on how inaccuracies in the QALY index

distort project selections and how averse he or she is to incorrect project choices as opposed to financial losses. Simulations indicate substitution of instruments distorts project orderings less than the adoption of additive structures if the QALY index need only order morbid states, but that the situation is reversed if the QALY must measure values (which applies depends on how the QALY is used to evaluate projects). The appropriate choice cannot be determined without reference to the researcher's cost situation and preferences.

2.4 Conclusion

One of the major challenges in health policy is the measurement of health status outcomes. QALYs provide measures which are free of the distortions and ethical biases inherent in health care markets, but they are not freely observable and must be obtained in experimental settings. QALYs assign appropriate values to health states (values that lead to policy decisions that maximize well-being) if the survey question used to derive the value is couched in the appropriate context for the type of policy being considered. Unfortunately, this increases the dimensionality of the problem facing the survey respondent and can dramatically increase the costs of obtaining QALY values, particularly if different types of policies are to be compared (e.g. those that affect length of life versus those that affect large groups of people). The second part of this chapter addressed the feasibility aspects of obtaining QALY values and verified a reconstruction method that can significantly reduce the number of health states that must be valued directly by survey respondents, but maintains the validity of the resultant QALY values.

This chapter suggests two avenues for further investigation. One is to carefully reassess the sorts of policy decisions which can be assisted with QALY-type data, and

the information content of these QALY values necessary for them to be valid in these applications. It may be that some types of policy analysis require more information content in the QALY measures than is feasible. The second avenue of research is to determine if reconstruction methods are possible over these extra dimensions of policy context (i.e. beyond morbidity). If this is the case, it should be feasible to use QALY data in a wide variety of health policy analyses.

Chapter 3

The Welfare Properties of Three Health Status Statistics

3.1 Introduction

A health status index is an index number which measures both the quantity (how long one lives) and quality (how well one is during this time) of health. It acts as an aggregator function of the various dimensions of health. This has two important implications. First, the various aspects of health must be valued relative to one another. Second, the index allows disparate health states to be compared.

Such statistics can help decision makers in three types of allocation decisions. First, one can determine which treatment path yields the best possible health outcome for a given patient. For instance, a patient with kidney disease may be treated with dialysis or organ transplantation. The health outcomes of these treatments are different (dialysis is very time consuming and leaves the patient lethargic; transplantation is followed by a variety of complications due to surgery and immunosuppressants and has a higher risk of immediate death). A health index can be used to compare these different outcomes and determine whether transplantation or dialysis yields “more” health.

Second, one may use a health status statistic to determine which patient should receive a given treatment (i.e. who benefits more from this treatment). This sort of decision confronts policy makers whenever treatments must be rationed. An example would be heart transplants, which are limited by the availability of donor hearts.

When one becomes available, a recipient must be selected from the waiting list. If policy makers wish to allocate resources to get the greatest possible level of health, then the patient with the greatest improvement in health, as measured by the health status index, should be chosen.

Third, one can use health status statistics to determine which programs generate the most health benefits for society and should be implemented (i.e. compare different treatments affecting different people). This involves comparisons of the health of groups of people. An example would be the decision to implement either a pancreatic transplant or a lung transplant program in a given region. The former improves the health of persons suffering from severe diabetes and the latter improves the health of persons suffering from severe respiratory disease. Assuming the costs of either program are equal, the decision will depend on how many people can be transplanted in each program with the given resources, the health improvements of the people treated, and how these improvements are measured.

Obviously, how a health status index measures the various aspects of a health state critically affects the types of decisions made above. The goal of this chapter is to assess three prominent health status indexes, human capital, willingness-to-pay, and QALYs (quality adjusted life years), to determine if decisions made on the basis of these measurements are consistent with individual and social welfare maximization. If the valuations inherent in these indexes are inconsistent with such objectives, policy makers may make decisions that leave society with a sub-optimal level of health.

3.1.1 Review of Health Status Measures

Three health status measures have been used by economists in health policy assessments: human capital (HK), willingness-to-pay (WTP), and the quality adjusted life year (QALY). These are described below.

Human Capital Measures

Human capital measures equate the value of a health state to the income that can be earned in that health state (usually expressed in present value terms).¹ For instance, if a person after kidney transplantation is only willing or able to work twenty hours a week at a non-strenuous job paying ten dollars per hour, the value of a year of life in this health state would be about ten thousand dollars. If the person lives five years in this state and the rate of discount is ten per cent a year, the value of this life would be about forty thousand dollars. The human capital measure of societal health is the level of G.N.P. (gross national product) generated under a given community profile of health status (i.e. the sum of all the individual human capital values).

The index is based on principles of national income accounting and measures the potential contribution to G.N.P. of an improvement in health status. The principle advantage of this statistic is that it is based on readily available income and work-force data. Its disadvantages are twofold. First, the statistic is inconsistent with Pareto optimality because the measure of marginal benefit ignores all non-financial aspects of well-being ² ³ (if transplantation made you feel better, but you could not find work because of discrimination by employers, the value of the health state with transplantation would be zero). Second, entitlement is based on income generating potential so that poor persons, particularly those not in the “formal” labour market and the retired, are discriminated against. The corporate executive is more likely to

¹Mishan (1976) describes both the traditional measure above and variations with either consumption or savings deducted (reflecting the position that only benefits to others and only benefits to the individual should matter respectively).

²This assumes the measure of marginal benefit is undistorted by imperfections in the labour market.

³Brent (1991) has devised a human capital model with time, valued at the wage rate, as the metric rather than income. This essentially equates the value of life with the value of full income, so the valuation problem remains. The conditions under which marginal income and marginal utility are synonymous were not identified.

get a heart transplant than the housewife.

Willingness-to-Pay Measures

Willingness-to-pay (WTP) measures health status changes in a standard revealed preference framework. The value of a health status change is the maximum amount of income the individual is willing to give up in order to secure the change. Assuming perfectly competitive markets for kidney transplants, fully informed patients and no insurance that distorts marginal prices, one could (hypothetically) find the value of a kidney transplant by observing the market price at which such transplants were traded. The usual aggregate statistic is the sum of the individual willingnesses-to-pay.

This index is based on principles of standard cost-benefit analysis. Because of various market imperfections, accurate willingness-to-pay values are difficult to achieve.⁴ It is obvious that willingness-to-pay is determined in part by ability to pay so that the wealthy are favoured over poor people in treatment, regardless of the source of their income (see Torrance [1986], Brent [1991]). There are serious consistency problems with the aggregate statistic, as has been shown by Boadway (1974) and Blackorby and Donaldson (1990). For instance, it is possible for the aggregate statistic to indicate that societal health is higher when a dialysis program replaces a kidney transplant program, and that societal health is higher when a transplant program replaces dialysis.

⁴WTP measures may be obtained through empirical observation and market data or by hypothetical survey questionnaires. Both approaches can be problematic since health care markets are badly distorted and do not span all relevant health characteristics, and surveys typically suffer from low response rates. These difficulties contributed to the search for alternative measures.

QALY Based Measures

QALYs were developed in response to concerns that health care resources should not be deployed by ability to pay ⁵(a fact reflected in numerous health care institutions), but on a basis still consistent with individual preferences. The QALY value is derived from preferences using a QALY instrument (see Chapter 2), a device by which the strength of preference between two (hypothetical) health states (described in a questionnaire) can be measured. The healthy year equivalent (HYE) is a special type of QALY where the duration of a health state in addition to morbidity is explicitly part of the health state description. In the kidney transplant example above, the individual would be described the morbidity associated with kidney transplants and told it would last, say, ten years. The individual would then be asked how many years in perfect health would leave him or her as well off as living for ten years with a transplant. This answer is the HYE value. The aggregate statistic is the arithmetic mean of the individual statistics (with a fixed population, this is equivalent to the sum of the individual statistics).

QALY values are more reliable than market based values because they are not influenced by market imperfections. They are not freely observable and can be expensive to obtain. Almost all attention has focused on the methods used to derive such a health status index (i.e. which instrument to use) and whether it is a utility number or not (see Chapter 2 and Butler [1990]). Hilden (1985) has questioned the interpersonal properties of QALYs because variations in the value of health relative to other goods are suppressed (Ware and Young [1979] provide evidence that such variations in the value of health exist). Butler (1990) has claimed QALYs are proportional to willingness-to-pay measures, but, like Hilden, does not demonstrate the

⁵Torrance et al. [1972] suggest that “economic” well-being should be held constant in the QALY survey, claiming this purges the measure of any ability to pay considerations.

conditions when this is so. While the arithmetic mean has been the only aggregate statistic used to date, “the assumptions which underlie any method of aggregation, and the policy implications of alternative methods, need to be more fully explored” (Loomes and McKenzie: 1989, p. 305).

Almost twenty years ago Torrance argued “the theory [of QALYs] should be extended to include multiple utility maximizing individuals, interpersonal comparisons of utility, and group decision-making” (1976a, p. 369). This review demonstrates the limited progress made since then. Furthermore, the performance of the QALY index has never been compared rigorously to the alternative health status indexes used by economists in the past. It is this assessment which is now undertaken.

3.1.2 Evaluation Criteria

The purpose of this chapter is to evaluate the performance of these indexes as measures of health status, not the decision statistics which sometimes incorporate them with other non-health information (such as the cost-utility ratio). Three principles form the basis of evaluation.

i) the individual is the best judge of his or her well-being

For the purpose of this paper, it is assumed well-being is synonymous with utility. Then this condition requires that the health status index assigns higher values to states more preferred by the individual, i.e. the index is an exact representation of the preference ordering. This position is adopted to prevent the policy maker from imposing treatments that the patient would not want. For instance, in the kidney example above, a health index may assign higher values to transplantation than dialysis. If the patient actually preferred dialysis, the policy maker who based his or her decision on the health index information would make the wrong choice.

While this position is gaining acceptance in the medical field, several controversial implications need to be identified. First, more paternalistic positions must be abandoned (one cannot say, as did Avorn (1984) and Harris (1987), that prevention of death should be the paramount concern unless the patient agrees that all quality of life should be sacrificed to extend life). Second, the extrinsic value of an individual's health (i.e. the value to others of his or her health) is deemed irrelevant. Finally, because the health status index measures health outcomes, not the processes by which they are achieved, it must be assumed that preferences are also defined over outcomes. People may think lotteries are a preferred selection mechanism for transplants over some medically determined criteria, but this is irrelevant since only the well-being associated with the health outcomes of transplantation count, not how they are achieved.

ii) all individuals are entitled to equal access to health benefits

Equal access to health care is a controversial principle, although it has often been promulgated by policy makers and occasionally manifests itself in health care institutions (e.g. social insurance). The reader should be mindful that other ethical positions exist (e.g. entitlement by "social worth" was once used to determine access to dialysis).

Equal entitlement requires that the value of a health status change (as measured by the health status index) be equal for all persons with the same preference ordering over health. If two people would rather have treatment than not, the policy maker should not be able to discriminate against one or the other in treatment allocation. For instance, suppose two people both prefer kidney transplantation to dialysis. One person is elderly and financially well-endowed, while the other is young and poor. Because both individuals prefer the same alternative, each should have an equal opportunity to receive a transplant. Given that the policy maker will allocate the donor

kidney according to which patient achieves the greatest benefit from the transplant over dialysis, equal entitlement requires that the health status improvement, as measured by the health status index, be the same for both individuals

iii) policy decisions should be guided by distributionally sensitive welfarist social preferences defined over the health states of the individuals in the community

Assume that social well-being may be represented by some set of preferences defined over community health profiles and this may in turn be represented by a social welfare function. Then this third principle implies social preferences must be defined over individual preferences (i.e. the social welfare function is a welfarist Bergson-Samuelson social welfare function), and that the aggregation of these individual preferences must reflect the degree of inequality aversion in the distribution of health outcomes across members of society held by the community (or a benevolent social planner). Welfarism is necessary so that the first and third principles do not contradict one another. Ethical flexibility in the distribution of health outcomes does not mean the second and third principles are inconsistent because the second is defined over identical health states and the third is defined over health states that differ.

3.2 Theoretical Framework for Assessment

In this section, each decision statistic is derived within a utility maximizing framework. This establishes the relationship of each decision statistic to individual well-being and allows the comparison of the welfare properties of each.

3.2.1 Individual's Optimization Problem

The individual seeks to maximize his or her well-being, which may be represented by an objective function defined over personal consumption and health status. Health

status affects utility directly (via general wellness) and indirectly through productivity (e.g. wages) and by constraining the amount of time that may be spent in any particular activity. Utility is assumed to accumulate over the lifetime.

Let $U : R^N \rightarrow R^1$ describe preferences over health (represented by morbidity, q , and a length of life, T), and other goods (described by the vector x). The individual maximizes utility by choosing the optimal allocation of time to various activities and income to various purchased goods subject to financial, institutional, and time use constraints. The individual's choices are constrained in three ways. It is assumed that the individual may allocate time to three activities: active leisure (t_1), passive leisure (t_2), and paid labour (L). Active leisure and labour time may be constrained by poor health (i.e. $t_1 \leq \tau_1(q)$, $L \leq \tau_L(q)$), but passive leisure is not, by definition, ever constrained. Time spent in all activities, including the labour market (L), must not exceed total time (T). Finally, the value of all purchases of x , valued at prices, p , must not exceed income earned (total time, T , less time spent at leisure activities, valued at the wage rate, $w(q)$) plus endowed income, I . Note that the wage rate is assumed to depend on health status. Health can then affect utility in three ways: through earning capacity, through time constraints, and directly.

Then the individual's problem is to find x , t_1 , and t_2 to solve:

$$\begin{aligned} \max_{x, t_1, t_2} \{ & U(x, t_1, t_2, q) : (i) \sum_i p_i x_i \leq w(q)L + I; \\ & (ii) \sum_{j=1}^2 t_j + L = T; (iii) t_1 \leq \tau_1(q); (iv) L \leq \tau_L(q) \}. \end{aligned} \quad (3.1)$$

(This problem can be restated in a dynamic context, with all variables time-scripted so that their evolution can be tracked:

$$\begin{aligned} \max_{x_1, \dots, x_T, t_{1,1}, \dots, t_{1,T}, t_{2,1}, \dots, t_{2,T}} \{ & U(x_1, \dots, x_T, t_{1,1}, \dots, t_{1,T}, t_{2,1}, \dots, t_{2,T}) : \sum_{k=1}^T \sum_{i=1}^I p_{ik} x_{ik} \leq \\ & \sum_{k=1}^T w_k(q_k)L_k + I; \sum_{j=1}^2 t_{jk} + L_k \leq 1 \forall k = 1, \dots, T; L_k \leq \tau_{Lk}(q_k); t_{1k} \leq \tau_{1k}(q_k) \forall k = \end{aligned}$$

$1, \dots, T\}$, where k denotes the time period, i the purchased commodity, and j the activity. Such a representation provides greater detail, although most of it is superfluous for the purposes here. The only critical difference comes in the calculation of the HYE function. This is discussed in the section where this statistic is derived.)

After solving this problem, optimal choices may be substituted into the utility function, U , to obtain an indirect utility function defined over given parameters. It is sometimes easier to derive the health statistics from a dual representation, but problems arise because of the presence of multiple constraints. The expenditure functions are not well-defined in these cases. There are two ways to remedy this situation.

The first is to view the indirect utility function as made up of a set of constrained indirect utility functions, one for each possible combination of binding constraints. There are four possible cases: neither labour nor leisure is constrained, active leisure is constrained, labour is constrained, and both labour and active leisure are constrained. If u is the realized value of utility at (x^*, t_1^*, t_2^*) (i.e. $u = U(x^*, t_1^*, t_2^*)$), then if time allocations are unconstrained:

$$u = V(p, w(q), w(q)T + I, q)$$

$$\equiv \max\{U(x, t_1, t_2) : \sum_i p_i x_i \leq w(q)(T - t_1 - t_2) + I\};$$

if $L = \tau_L(q)$ (i.e. labour is constrained):

$$u = \hat{V}(p, w(q), w(q)\tau_L(q) + I, T - \tau_L(q), q)$$

$$\equiv \max\{U(x, t_1, t_2) : \sum_i p_i x_i \leq w(q)\tau_L(q) + I; t_1 + t_2 = T - \tau_L\};$$

if $t_1 = \tau_1(q)$ (i.e. active leisure is constrained):

$$u = \tilde{V}(p, w(q), w(q)(T - \tau_1(q)) + I, \tau_1(q), q)$$

$$\equiv \max\{U(x, t_1, t_2) : \sum_i p_i x_i \leq w(q)(T - \tau_1 - t_2); t_1 = \tau_1\};$$

if $L = \tau_L(q), t_1 = \tau_1(q)$ (i.e. both active leisure and labour are constrained):

$$u = \bar{V}(p, w(q), w(q)\tau_L(q) + I, T - \tau_L(q) - \tau_1(q), \tau_1(q), q)$$

$$\equiv \max\{U(x, t_1, t_2) : \sum_i p_i x_i \leq w(q)\tau_L(q) + I; t_2 = T - \tau_L - \tau_1\}. \quad (3.2)$$

For each constrained indirect utility function there exists one well-defined constraint (the time use constraints which bind in each case may be substituted into the income constraint to achieve a single binding constraint). Then for each constrained indirect utility function there exists a well-defined expenditure or cost function which is dual to it, and this may be recovered by inverting the constrained indirect utility function. This yields a set of expenditure functions:

if time use is not constrained:

$$w(q)T + I = E(u, p, w(q), q)$$

$$\equiv \min\{\sum_i p_i x_i + w(q)(t_1 + t_2) : U(x, t_1, t_2) \geq u; t_1 + t_2 + L = T\};$$

if $L = \tau_L(q)$:

$$w(q)\tau_L(q) + I = \hat{E}(u, p, w(q), T - \tau_L(q), q)$$

$$\equiv \min\{\sum_i p_i x_i + w(q)(t_1 + t_2) : U(x, t_1, t_2) \geq u; t_1 + t_2 = T - \tau_L\};$$

if $t_1 = \tau_1(q)$:

$$w(q)(T - \tau_1(q)) + I = \tilde{E}(u, p, w(q), \tau_1(q), q)$$

$$\equiv \min\{\sum_i p_i x_i + w(q)(t_1 + t_2) : U(x, t_1, t_2) \geq u; t_1 = \tau_1; t_1 + t_2 + L = T\};$$

if $L = \tau_L(q), t_1 = \tau_1(q)$:

$$w(q)\tau_L(q) + I = \hat{\tilde{E}}(u, p, w(q), T - \tau_L(q) - \tau_1(q), \tau_1(q), q)$$

$$\equiv \min \left\{ \sum_i p_i x_i + w(q)(t_1 + t_2) : U(x, t_1, t_2) \geq u; t_1 = \tau_1; t_1 + t_2 + \tau_L = T \right\}. \quad (3.3)$$

While these functions are well-defined, they are restrictive to work with in cases where the set of binding constraints changes.

The second method is to endogenize prices so that the time use constraints never bind and the individual in effect faces one “composite” budget constraint. In this case the individual’s problem becomes:

$$\max_{x, t_1, t_2} \{ U(x, t_1, t_2, q) : (i) \sum_i p_i x_i + w_1 t_1 + w_2 t_2 \leq w(q)L + I + w_1 t_1 + w_2 t_2 \}, \quad (3.4)$$

where w_1 and w_2 are shadow prices on the two leisure activities (see Sen [1972] for a discussion of shadow pricing constrained goods). If labour is constrained at market prices, the individual must be induced to consume more leisure (until the labour constraint is no longer binding), which is accomplished by reducing the (shadow) price of both leisure activities below the market wage ($w_1 = w_2 < w(q)$). If active leisure is constrained, the individual must be induced to consume less active leisure (until the constraint is no longer binding), which is accomplished by raising the (shadow) price of active leisure above the market wage rate ($w_1 > w(q) = w_2$). Finally, if both active leisure and labour are constrained, the individual must be induced to consume more passive leisure and less active leisure by reducing the shadow price of passive leisure and increasing the shadow price of active leisure ($w_2 < w(q) < w_1$). Obviously, the shadow prices depend on the market wages and prices, the levels of the constraints and endowments, and preferences.

Solving the individual’s problem and substituting the optimal allocations into the utility function, it is possible to obtain the global indirect utility function (the solution to equation (3.1)):

$$u = V(p, w(q), w(q)G(T, q, \tau_1(q), \tau_L(q)) + I, H(T, q, \tau_1(q), \tau_L(q)), q) = \mathcal{V}(p, q, T, I) \quad (3.5)$$

where $G(T, q, \tau_1(q), \tau_L(q))$ and $H(T, q, \tau_1(q), \tau_L(q))$ are functions of the shadow prices, endowments and constraints that indicate the full income and the leisure allocations possible.

3.2.2 Derivation of the Health Status Indexes

A health care project is described by its effect on health status (which changes from (q^B, T^B) to (q^A, T^A) , where A denotes “after” and B “before” the project). Associated with these two states are utility levels u^A and u^B . It is assumed costs of the project are covered by a third party payer and do not, directly or indirectly, affect the individual’s situation.

3.2.3 Human Capital Measures (HK)

Recall that the human capital measure equates the value of a health state with the income that can be earned in that health state. This requires that the amount of labour effort in a particular health state supplied by the individual be identified, as well as the return to this labour effort (i.e. the wage rate). To do this, the set of constrained expenditure functions is employed, and each case is assessed separately.

Unconstrained Case

To find the labour supply in this case, apply Shepard’s lemma to the unconstrained expenditure function:

$$L^*(u, p, w(q), q) = \frac{\partial E(u, p, w(q), q)}{\partial w(q)}. \quad (3.6)$$

A health status improvement is measured by the change in earned income (ΔHK) arising from the health status change. Given the labour supply function above, and

assuming labour and active leisure use are unconstrained in both the before and after states,⁶ this may be expressed as

$$\Delta HK = w(q^A) \frac{\partial E(u^A, p, w(q^A), q^A)}{\partial w(q^A)} - w(q^B) \frac{\partial E(u^B, p, w(q^B), q^B)}{\partial w(q^B)}. \quad (3.7)$$

Labour Constrained Case

If labour choice is constrained, then labour can only be provided to the level of the constraint and

$$HK = w(q)\tau_L(q), \quad (3.8)$$

such that the health status change is measured as

$$\Delta HK = w(q^A)\tau_L(q^A) - w(q^B)\tau_L(q^B). \quad (3.9)$$

Leisure Constrained Case

The individual's labour choice incorporates any binding leisure restrictions. This may be recovered using Shepard's lemma on the leisure constrained expenditure function:

$$L^* = \frac{\partial \tilde{E}(u, p, w(q), \tau_1(q), q)}{\partial w(q)}. \quad (3.10)$$

Thus, the value of a health status change may be expressed as

$$\begin{aligned} \Delta HK = & w(q^A) \frac{\partial \tilde{E}(u^A, p, w(q^A), \tau_1(q^A), q^A)}{\partial w(q^A)} \\ & - w(q^B) \frac{\partial \tilde{E}(u^B, p, w(q^B), \tau_1(q^B), q^B)}{\partial w(q^B)}. \end{aligned} \quad (3.11)$$

⁶This assumption is made for convenience. The human capital function is still well defined if the after state is characterized by one set of constraints, and the before state by a different set. One would have to substitute the appropriate labour supply function according to the constraints binding in that case.

Labour and Leisure are Constrained

If both constraints are binding, the amount of labour supplied equals the labour constraint and the labour constrained case holds.

3.2.4 Willingness-to-Pay

The willingness-to-pay (cv) is the maximum amount of income the individual is willing to give up in order to have the change in health status (the compensating variation, rather than the equivalent variation, is adopted since health care projects typically involve moving to better states that people are willing to pay for – the “do no harm” ethic ensures few converse situations arise). This is the amount of income that can be deducted in the after state and leave the individual as well off as in the before state.

Then, recalling equation (3.5), cv may be defined implicitly:

$$\mathcal{V}(p, q^A, T^A, I - cv) = \mathcal{V}(p, q^B, T^B, I) = u^B. \quad (3.12)$$

3.2.5 QALY/HYE Measures

In this analysis, the HYE (healthy year equivalent) is employed rather than a QALY based on an arbitrary time frame. This is done so that subsequent welfare results are not conditioned on measurement error. It is also assumed that the individual accounts for the effects on his or her income that result from a change in health status. This contradicts the position taken by Torrance et al. (1972) that the QALY be purged of all income effects (they suggest using an income replacement scheme so that there are no financial effects to the respondent – or his or her family – from the health status change). The position adopted by Torrance et al. implicitly assumes all financial effects are captured in the cost assessment, rather than in the health outcome assessment (e.g. increases in earnings would be counted as cost savings). In reality,

the cost assessment only incorporates costs incurred by the funding agency (e.g. the Ministry of Health). This is reasonable when one considers the funding agency must choose programs to maximize health subject to its own budget constraint. If the cost assessment included costs for which the agency was not responsible or cost savings which it could not recoup (such as income effects of the patients), the choices made would either not exhaust or exceed the budget constraint.

The HYE is similar to the willingness-to-pay measure since both are based on trade-offs that achieve welfare equivalence. However, the numeraire for HYE is time, not income, and the reference state for HYE is fixed at perfect health and not some arbitrary “after” state.

In the HYE assessment, the individual gives up time in perfect health until he or she is indifferent between that state and the one described, i.e.

$$\mathcal{V}(p, \bar{q}, (T - m), I) = \mathcal{V}(p, q, T, I) = u. \quad (3.13)$$

The HYE value for a project is the difference between the HYE for the before and after states

$$\Delta HYE = (T^A - m(q^A)) - (T^B - m(q^B)). \quad (3.14)$$

Note that in the unconstrained case the change in healthy year equivalent and the willingness-to-pay measures are proportionately related by the market wage rate.

(In the dynamic context discussed before, time foregone must be valued at an average wage rate, i.e. in the unconstrained case: $\bar{w}(T - m) = \sum_{t=1}^{T-m} w_t(\bar{q}_t)$. Then $T - m = \frac{1}{\bar{w}}(E(u, p_t, w_t(\bar{q}_t), \bar{q}_t) - I)$. If the wage schedule at perfect health varies over time, then \bar{w} varies as the time frame varies. Different health states are then valued by different functions of utility, causing potential problems with exactness. However, if the wage schedule is relatively continuous and health states are not too disparate, a reasonable approximation is achieved.)

3.3 Choosing Treatments for a Given Individual

This section examines the welfare properties of the three health status indexes when they are used to determine the “best” treatment for an individual. Recall that the efficiency criteria dictate that the appropriate choice is the one which assigns the individual to his or her most preferred outcome state. This requires that the indicator be consistent (exact) with the individual’s preference ordering over outcomes.

A *health status index* is said to be exact if it assigns higher values to health states that are more preferred by the individual in question. A health state is completely described by two characteristics: morbidity (q), and length of life (T). By definition, the utility function, U , is exact.

3.3.1 Exactness of the Health Status Index

Because a health status change affects only q and T , is it possible to express the utility function over different health states as $\bar{v}(q, T) = \mathcal{V}(p, q, T, I)$ (where p and I are fixed variables). Then a health status index, HS, is exact if and only if

$$HS(q, T) = \phi(\mathcal{V}(p, q, T, I)) = \phi(\bar{v}(q, T)) \quad (3.15)$$

(ϕ is some increasing monotonic transform).

Human Capital Measures

Human capital measures are not based directly on preferences, but on one behavioral manifestation of these preferences:

$$HK = w(q)L^*, \quad (3.16)$$

where the labour supply, L^* , usually depends on the utility level (in the labour constrained case, the labour supply depends on the constraint only).

Lemma 1 (HK): When labour supply is unconstrained, HK is exact (i.e. assigns higher values to more preferred states) if and only if the health care project affects T only and labour is a normal good. Labour constrained HK is exact if and only if the health care project affects only the labour constraint.

Proof: see Appendix E.

Note that the leisure constrained case is identical to the unconstrained case, except for the additional argument in the expenditure function, $\tau_1(q)$. It, like $w(q)$, must be the same for all projects compared. Because of this, the human capital measure is never exact between states of the world characterized by different binding time use constraints.

It is clear that the human capital statistic is not exact for most health state comparisons. It cannot compare two states where both morbidity and mortality change, nor can it compare two states characterized by different binding time use constraints. Even for the very narrow set of health states which it potentially can evaluate, it is exact only for very restricted (and unlikely) preferences. Hence, it is not a suitable index for evaluating health states at the individual level.

Willingness-to-Pay Measures

Unlike the human capital measure, the willingness-to-pay (WTP) measure is based on a direct comparison of well-being associated with any two states.

Lemma 1 (WTP): WTP is exact (i.e. takes on positive values when the “after” state is preferred to the “before” state, and vice versa).

Proof: see Appendix E.

Note that the presence of constraints has no effect on the exactness properties of the WTP measure.

Healthy Year Equivalents

The healthy year equivalent (HYE) is similar to WTP, although the valuation of health states depends on the value of time, rather than income.

Lemma 1 (HYE): HYE is exact (i.e. it assigns higher values to more preferred states).

Proof: see Appendix E.

Note that the presence of time use constraints has the same effect on HYE as on WTP.

3.3.2 Discussion

Exactness results are summarized in Table 3.1. HK statistics cannot compare projects involving different levels of morbidity, and can only rank projects involving different lengths of life if labour is a normal good (implying at least one type of leisure is inferior) and the set of binding time use constraints is constant. Willingness-to-pay measures and healthy year equivalents, on the other hand, are always exact.

In conclusion, the human capital measures are the least satisfactory in this type of decision framework (single individual). They cannot evaluate health states with different degrees of morbidity, and are consistent over length of life only under restricted (and unlikely) preferences. It is likely that the wrong treatments will be chosen if treatment choices are made according to the rankings of the human capital index. The willingness-to-pay measure is exact, its only disadvantage being the difficulty in obtaining accurate data. The QALY based measures are also exact, and may be more reliable than the willingness-to-pay measures. Theoretically, both indexes indicate appropriate treatment paths, although the willingness-to-pay measures may be biased because of market distortions. These distortions have to be assessed in light

of the significant costs of acquiring the alternate QALY data.

	q	T
HK	inexact	$L_I > 0$
cv	exact	exact
HYE	exact	exact

Table 3.1: Conditions for Exactness

3.4 Choices Between Individuals

In this section, attention is turned from decisions over which treatment to give to a particular individual to decisions about which individual should receive a particular treatment (e.g. who should receive a heart transplant when donor hearts are scarce). Decisions of this nature involve the type of allocative issues which led to dissatisfaction with both the willingness-to-pay and human capital approaches and the development of QALY based alternatives.

It is assumed that the policy maker chooses to allocate treatments to the people who benefit the most by them (as measured by the health status index). For instance, a heart transplant allows person A to return to work as a corporate executive making 100,000 dollars a year, or person B to continue to live on a pension in retirement. Not receiving a transplant results in death for either person. The human capital statistic measures the annual benefit to person A as 100,000 dollars more than the benefits to person B. The policy maker would then choose person A over person B for the next donor heart because this choice results in the greater measured health benefit.

The question is whether such decisions are consistent with the principle of equal entitlement.⁷ This requires that all persons who prefer one treatment path over

⁷Other ethical positions are possible (e.g. merit based allocations), although equal entitlement is fairly common. What is important is to recognize that some such ethical judgement is inherent in this type of decision and that these judgements should be made explicitly. Acceptance of arbitrary

another have an equal opportunity to receive that treatment (i.e. the health status statistic measures the benefits of such a treatment as equal for all persons who want this treatment). This is done in two steps. First, it is determined if such health status measures vary across individuals with identical preferences but different endowments. Second, the variations in health status measures across individuals with identical endowments but different preferences are assessed.

3.4.1 Variations Due to Endowments

In this section, each health status statistic is evaluated for independence from (i) income (wealth and earnings), (ii) health (life expectancy and co-morbid effects), and (iii) time use constraints. Because choices between individuals are based on measured differences in health status as a result of treatment, the appropriate function for analysis is the difference in the health status statistic evaluated at the treatment and no treatment states.

Human Capital Measures

The human capital statistic measures the value of a health status change (denoted by the move from (q^B, T^B) to (q^A, T^A)) as the change in earned income due to the health status change:

$$\Delta HK = w(q^A)L^*(q^A, T^A) - w(q^B)L^*(q^B, T^B), \quad (3.17)$$

where $L^*(q, T)$ is the supply of labour in health state (q, T) .

income effects

The (unconstrained labour supply) human capital statistic is invariant to the level of wealth (I) if and only if the income elasticity of labour supply is perfectly elastic,

value standards in morally inept.

and is invariant to earning capacity (w) if and only if the price elasticity of labour supply is unitary.⁸

It is relatively easy to derive these results by differentiating equation (3.17) (with the appropriate unconstrained labour supply functions substituted into the equation) by I and w respectively. In the former case, $\partial\Delta HK/\partial I = 0$ (the statistic is independent of wealth) if and only if $-dw/(w(q^A) + dw) = d(\frac{\partial L^A}{\partial I})/\frac{\partial L^A}{\partial I}$ (where 'd' denotes the change between the before and after states and L^A is the labour supply in the after state). This must apply to all projects, including those that affect only length of life. But then $dw = 0$, so independence requires $d(\frac{\partial L^A}{\partial I}) = 0$ (i.e. the income elasticity must be the same for all states). Imposing this condition when the project affects wages, the independence condition becomes $dw(\partial L/\partial I) = 0$. Since $dw \neq 0$, this requires that the income elasticity of labour supply be zero.

Assume that wages differ between individuals by a constant amount. Then the effect of a change in wages may be found by differentiating equation (3.17) with respect to this factor. Then $\partial\Delta HK/\partial w = 0$ if and only if $L^*(\epsilon_{Lw} + 1) = c$ (where ϵ_{Lw} is the price elasticity of labour supply, L^*) for all states. Unless the case where $L^* = 0$ can be eliminated *a priori*, this can only hold if $c = 0$. Then independence requires $\epsilon_{Lw} = -1$ (unitary price elasticity).

Consider the effects when these elasticities do not hold. Suppose the income elasticity of labour is negative (people with large wealth holdings work less). A heart transplant allows a person to live and labour activity is not constrained after the transplant. Consider two people, one with large wealth holdings and one with little wealth, but otherwise identical. Then, because the income elasticity of labour is negative, the individual with small wealth holdings spends a larger portion of this

⁸In the labour constrained case, the labour supply is invariant to the level of income or wages, so the conditions do not hold.

extended life working than the wealthier individual so that the human capital statistic favours poor individuals when the donor heart is allocated.

health effects

The human capital statistic is invariant to life expectancy (T) if and only if the supply of labour is invariant to the length of life remaining and is invariant to co-morbid effects if and only if the change in earnings is the same for all health states.

These results can be found by the same methods above. Differentiating equation (3.17) with respect to T and following the same steps as in the wealth case, independence from life expectancy requires $\partial L^*/\partial T = 0$. Co-morbid effects can also be found by differentiating (3.17) with respect to q_j , some morbid characteristic unaffected by the health status change. This yields the independence condition: $\partial \Delta HK / \partial q_j = w(q^A) \partial L^A / \partial q_j - w(q^B) \partial L^B / \partial q_j + (\partial w(q^A) / \partial q_j) L^A - (\partial w(q^B) / \partial q_j) L^B$.

In general, people with more years to live will work more ($\partial L^*/\partial T > 0$), with the exception of people who are constrained in labour time. Then, the human capital statistic discriminates against the elderly (e.g. a young person will get an organ transplant over an older person because, *ceteris paribus*, the period in which income can be earned is longer for the youngster). The effects of co-morbidity are more difficult to predict: a pre-existing disability may either inflate or dampen the earnings gain from the health status improvement. It is not possible to predict if this statistic discriminates against the disabled.

time use constraints

To examine the effects of time use constraints, the constrained versions of labour supply are used. It is assumed that only one set of constraints is binding in both the before and after treatment states. This allows problems with non-linear constraints to be circumvented, while allowing comparisons across persons who are differentially constrained.

Generally, human capital measures are independent of time use constraints if and only if labour supply is unaffected by such constraints. In the labour constrained case, this requires that the labour constraint not be binding.

If the labour constraint is binding (regardless of whether or not the leisure constraint is binding or not), the effect on the human capital measure may be found by differentiating the labour constrained case by the labour constraint: $\partial \Delta HK / \partial \tau_L = (w(q^A) - w(q^B)) \partial L^* / \partial \tau_L$. Given that wages cannot be restricted *a priori* to be equal, this requires that $\partial L^* / \partial \tau_L = 0$ (i.e. the labour constraint does not bind). Similarly, leisure constraint effects can be found by differentiating the leisure constrained labour functions: $\partial \Delta HK / \partial \tau_1 = w(q^A)(\partial L^A / \partial \tau_1) - w(q^B)(\partial L^B / \partial \tau_1)$.

While it is clear labour constrained individuals (the mandatorily retired) are discriminated against when the human capital statistic is used to allocate resources, the leisure constrained individual may or may not be.

Willingness-to-Pay Measures

The willingness-to-pay statistic may be represented in a global indirect utility function which encompasses all possible constraints (multiple constraints are a problem for the dual function only):

$$\begin{aligned} & \mathbf{V}(p, w(q^A), w(q^A)G(q^A, T^A, \tau_1(q^A), \tau_L(q^A)) + I - cv, H(q^A, T^A, \tau_1(q^A), \tau_L(q^A)), q^A) \\ &= \mathbf{V}(p, w(q^B), w(q^B)G(q^B, T^B, \tau_1(q^B), \tau_L(q^B)) + I, H(q^B, T^B, \tau_1(q^B), \tau_L(q^B)), q^B). \end{aligned} \quad (3.18)$$

This function is used to derive several of the independence conditions below because the results obtained can be generalized beyond the time use unconstrained case. To compare the effects of time use constraints, however, the segmented constrained indirect utility functions are used. This is allowable because it is the difference in

willingness-to-pay for a given health status change between the case when the same time use constraints bind in both the before and after states, and when they do not bind in either state that is of interest.

income effects

The willingness-to-pay measure is independent of wealth (I) if and only if income is additively separable from health status (q and T) in the utility function, and is independent of earning potential if the effect on the marginal utility of income of a change in wages is proportional to the effect on the marginal utility of health status.

To demonstrate the above, express equation (3.18) in its reduced form (with all constants suppressed) to obtain the equivalent expression $V(q^A, T^A, I - cv) = V(q^B, T^B, I)$. Then cv is independent of I if and only if it is the same for all values of I . If cv is independent of I , then $cv = f(q^A, T^A, q^B, T^B)$. Set $(q^A, T^A) = (q, T)$ and $(q^B, T^B) = (\bar{q}, \bar{T})$ (i.e. the before state is fixed at (\bar{q}, \bar{T})). Then the fixed values may be suppressed in the expression for cv : $cv = \tilde{f}(q, T)$. Substituting into the indirect utility function, one obtains: $V(q, T, I - cv) = V(\bar{q}, \bar{T}, I)$. Given that cv is independent of I , choose I to be $I = cv + m$, where m can take on any value. Then the indirect utility function becomes $V(q, T, m) = V(\bar{q}, \bar{T}, cv + m)$ which, because (\bar{q}, \bar{T}) are fixed, may be expressed $V(q, T, m) = v(cv + m) \stackrel{o}{=} cv + m$. But $cv = f(q, T)$, so $V(q, T, m) \stackrel{o}{=} f(q, T) + m$, which is the desired result. If $V(q, T, I) \stackrel{o}{=} f(q, T) + I$, then $cv = f(q^A, T^A) - f(q^B, T^B)$, which is independent of I .

To derive the earnings result, assume that $dw(q^A) = dw(q^B)$, and that the change in health status is small (so that cv may be defined as a function of derivatives⁹). Expressing (3.18) in the reduced form, independence requires $V(q^A, T^A, I - cv, w(q^A) + dw) = V(q^B, T^B, I, w(q^B) + dw)$ for all dw . Solving for cv and differentiating with

⁹The results are easily extended to large changes in health status since any discrete change in health status may be expressed as a series of small changes. Given that these conditions hold over all small changes, they must then hold for any large change.

respect to dw , one obtains $\partial cv/\partial w = \partial(\frac{V_{HS}}{V_I})/\partial w$ (where $V_{HS}dHS = V_qdq + V_TdT$, and subscripts denote derivatives), and $\partial cv/\partial w = 0$ if and only if $V_{HS,w}/V_{HS} = V_{I,w}/V_I$.

In both cases, the effect of a change in income on willingness-to-pay depends on how that change affects the marginal value of health relative to the marginal value of income. For the statistic to be biased to the rich, the marginal value of health must increase more than proportionately than the marginal value of income when wealth or wages change. In the case of wealth, it is generally accepted that the marginal value of income falls as wealth increases, and evidence from Viscusi and Evans (1990) suggests the marginal value of morbidity increases with income. Thus, the measure is biased against the people with small wealth holdings. It is not clear that the same bias holds for earning potential, however.

health effects

The willingness-to-pay measure is independent of life expectancy (T) or co-morbid effects (q_j) if and only if each element of health status is weakly separable from all other elements.

This result can be demonstrated using either T or q_j (q_j is used here). Assume that the change in health status is very small (so that derivatives may be used). Then, since cv may be expressed $cv = \frac{V_T}{V_I}dT + \sum_i \frac{V_{q_i}}{V_I}dq_i$, independence requires $\partial cv/\partial q_j = 0$, or $\partial(\frac{V_T}{V_I})/\partial q_j = \partial(\frac{V_{q_i}}{V_I})/\partial q_j = 0$ for all q_i . But this is the definition of weak separability of q_j from all other elements of health and income. Since this must hold for every aspect of health, this implies all elements must be weakly separable from each other, i.e. $V(q_i, q_j, T, I) = \bar{V}(v_1(q_i), v_2(q_j), v_3(T), v_4(I))$.

Again, the effects of a change in health status unrelated to treatment on willingness-to-pay depend on the relative effects on the marginal value of income and health. Evidence from Viscusi and Evans (1990) suggests the marginal utility of income is increasing in health, and standard economic theory suggests the marginal

utility of any one good (e.g. health) should be subject to diminishing returns. In this case, the marginal utility of health actually falls as the marginal utility of income increases, suggesting the aged and disabled are, *ceteris paribus*, favoured in treatment selection when this measure is used.

time use constraints

Time use constraints (either labour or leisure) do not affect willingness-to-pay if they are weakly separable from health and income in the utility function.

The purpose of this section is to compare the willingness-to-pay of an individual whose time is always constrained against one whose time is never constrained. Thus, the segmented constrained indirect utility functions are employed since only one set of constraints is ever binding in either case. In the unconstrained case, denoted by “ u ”, $cv^u = E(u_u^A, p, w(q^A), q^A) - E(u_u^B, p, w(q^A), q^A) = \bar{E}^u(u_u^A) - \bar{E}^u(u_u^B)$. The constrained case, denoted by “ c ”, may be expressed in similar fashion as $cv^c = \bar{E}^c(u_c^A) - \bar{E}^c(u_c^B)$. Then the difference in the willingness-to-pay due to the constraint is $cv^u - cv^c$. This is equal to zero if and only if $\bar{E}^u(u_u^A) - \bar{E}^u(u_u^B) = \bar{E}^c(u_c^A) - \bar{E}^c(u_c^B)$, or $(\partial \bar{E}^u / \partial u) du^u = (\partial \bar{E}^c / \partial u) du^c$. Since $\partial E / \partial u = \frac{1}{\partial V / \partial I}$, and $du = \partial V / \partial HS$ (where $dHS = \sum_i dq_i + dT$), this implies $MRS_{HS,I}^u = MRS_{HS,I}^c$ (i.e. the $MRS_{HS,I}$ is unaffected by the presence of the constraint, which defines weak separability).

Not surprisingly, the condition for independence of willingness-to-pay from a constraint is very similar to the condition for any other factor. If, for instance, the presence of a labour constraint makes the marginal value of income increase relative to the marginal value of health (as would be expected for improvements in T), the willingness-to-pay measure will favour the unconstrained individual (e.g. the individual in mandatory retirement is less likely than someone who may choose to work to be selected for an organ transplant).

Healthy Year Equivalents

As with the willingness-to-pay measure, the healthy year equivalent can be implicitly defined in a global utility function. Results derived from these functions can be generalized to all sets of constraints. The HYE may be expressed:

$$\begin{aligned} & V(p, w(q^A, w(q^A)G(q^A, T^A, \tau_1(q^A), \tau_L(q^A)) + I, H(q^A, T^A, \tau_1(q^A), \tau_L(q^A)), q^A) \\ & = V(p, w(\bar{q}, w(\bar{q})G(\bar{q}, T^A - m, \tau_1(\bar{q}), \tau_L(\bar{q})) + I, H(\bar{q}, T^A - m, \tau_1(\bar{q}), \tau_L(\bar{q})), \bar{q}). \end{aligned} \quad (3.19)$$

Then $\Delta HYE = (T^A - m(q^A, T^A) - T^B + m(q^B, T^B))$. When constraints themselves are assessed, the segmented constrained indirect utility function is used for the reasons given in the willingness-to-pay section.

income effects

Wealth does not affect HYE's if and only if wealth is separable from health (q and T). HYE's are independent of earning capacity (wages) if and only if the effect on the marginal utility of morbidity is proportional to the effect on the marginal utility of longevity.

These results are similar to those in the willingness-to-pay case, although their derivation is somewhat more involved. Differentiating the ΔHYE function with respect to income, one obtains $\partial \Delta HYE / \partial I = \partial HYE(q^A, T^A) / \partial I - \partial HYE(q^B, T^B) / \partial I$. Then $\partial \Delta HYE / \partial I = 0$ if and only if $\partial HYE(q, T) / \partial I = c$ for all (q, T) . But since $\partial HYE(\bar{q}, T) / \partial I = 0$ (perfect health is assigned a value of T regardless of the level of wealth), $\partial HYE(q, T) / \partial I = 0$ for all (q, T) . Given $HYE(q, T) = MRS_{q,T}(q - \bar{q})$, (again, assuming the change in q is sufficiently small to express HYE explicitly), $\partial HYE / \partial I = 0$ if and only if $\partial MRS_{q,T} / \partial I = 0$, which defines weak separability (because any large change in health status may be viewed as a series of

small changes, and because this condition holds at every point, then this condition must hold for all changes in health status regardless of size). Similarly, independence from earning capacity requires $\partial HYE(q, T)/\partial w = 0$ for all (q, T) . Using $HYE(q, T) = (V_q/V_T)(q - \bar{q})$, this requires $V_{qw}/V_q = V_{Tw}/V_T$.

The HYE favours the wealthy if an increase in wealth (or wage) increases the marginal utility of morbidity relative to the marginal utility of time. *A priori*, it is not clear whether this is the case or not. However, if time use is unconstrained, it is the case that $V_I = wV_T$. If this substitution is made in the independence expression for wealth, the condition for independence and the direction of bias is the same as the willingness-to-pay case ($\partial \Delta HYE/\partial I = w\partial(V_q/V_I)/\partial I$) and the wealthy are probably favoured by this statistic. This proportionality between the HYE and the willingness-to-pay statistics is lost if time use is constrained.

health effects

The HYE is independent of the length of life (T) if and only if length of life is additively separable in the utility function, whereas it is independent of co-morbid effects if and only if each element of morbidity is weakly separable from all the other elements.

To derive these effects, apply the same arguments as in the income case to obtain $\partial HYE(q, T)/\partial T = c$ for all (q, T) . But $\partial HYE(\bar{q}, T)/\partial T = 1$, since a year in perfect health is worth a year in perfect health. Thus, $\partial HYE(q, T)/\partial T = 1$ for all (q, T) . This requires $V(q, \lambda T) = V(\bar{q}, \lambda T - m)$ for all $\lambda > 0$. Let $\lambda = 1/T$. Then $m = \bar{v}(q)$. Since $T - m \stackrel{\circ}{=} U(q, T)$, this implies $U(q, T) \stackrel{\circ}{=} \bar{v}(q) + T$. Sufficiency is obvious. To derive the results for co-morbid effects, repeat the exercise used in the willingness-to-pay case with T substituted for I .

These results suggest the aged are discriminated against if the marginal utility of time falls more than the marginal utility of morbidity as time increases. There is

some empirical evidence to suggest that this is the case (see Brooks [1986] for evidence that QALY values are concave with respect to time). This is the HYE ability to pay analogue. The disabled are discriminated against if the disability reduces the marginal value of the morbid state proportionately less than the marginal value of length of life. Because the value of an extra year is reduced by a disability, it is possible that the HYE discriminates against the disabled, especially if the different components of morbidity are substitutes (e.g. the loss of one faculty, such as vision, makes one more dependent on other faculties, such as hearing).

time use constraints

Time use constraints (either labour or leisure) do not affect HYE's if the constraints are weakly separable from health and income in utility.

This result may be derived by the same methods used in the willingness-to-pay case. Thus, one would expect the retired individual to be discriminated against by the HYE or either of the other two health status statistics.

Summary

The majority of the conditions for independence derived above are inconsistent with observed preferences. Where possible, empirical evidence is used to determine the most probable direction of bias of the three health statistics. Not surprisingly, the human capital measure is biased towards individuals who are more inclined or able to earn labour income (the people with small financial reserves, those who earn the greatest return on work effort, the young and those who are allowed to work). It is not clear what the direction of bias is for those who are disabled or otherwise constrained in their leisure activities because these factors may increase or decrease the incentives to work. The willingness-to-pay measure is probably biased to the wealthy (an ability to pay argument), and surprisingly, the aged and disabled (because the

marginal utility of income has been found to be positively correlated with health, as a person becomes healthier, he or she is less inclined to sacrifice wealth for additional health). While it is clear the retired are also discriminated against, the effects of wage differentials and leisure constraints are not so apparent. Finally, it seems likely that the young and healthy are favoured by the HYE statistic, as well as those who are not labour constrained. It is not clear how the HYE is affected by wealth or earnings, although the wealthy are favoured if time use is unconstrained.

In conclusion, it is clear the person in mandatory retirement is discriminated against by all three statistics. The elderly and disabled are the most favourably treated by the willingness-to-pay statistic, while they may be least favourably treated by the HYE. The people with substantial wealth holdings are favoured the most by the willingness-to-pay statistic (and maybe the HYE) and the least by the human capital statistic.

3.4.2 Preference Variation

In this section, the implications of preference variation are assessed. It is assumed everyone has identical endowments, including time use constraints, and the same preference ordering over the two health states under consideration (this assumption restricts the marginal rate of substitution between morbidity and longevity to be the same among all individuals and imposes weak separability of health from non-health goods in the utility function). Because the preference ordering over the sub-space of health is restricted to be the same for all agents, equal entitlement requires the health status measures be invariant to any other aspect of preferences.

Human Capital Measures

Because the human capital measure is an inexact measure of health, it is not possible to relate it to the marginal rate of substitution between morbidity and length of life. It is obvious that it discriminates against people who, at the margin, value the consumption of leisure more than purchased goods (as reflected in the value of income), since these people, *ceteris paribus*, work less.

Willingness-to-Pay Measures

Assume that the health status change is very small (see the previous section for the sufficiency argument that any condition derived for small changes also holds for large changes). Then the willingness-to-pay measure is directly related to the marginal rate of substitution between any element of health and income (and, by transitivity, is consistent with the marginal rate of substitution between any two elements of health), evaluated at the “after” health state. Given that $cv = \frac{V_{HS}}{V_I} dHS$ (where $V_{HS}dHS = V_TdT + \sum_i V_{q_i}dq_i$), it is clearly inversely related to the marginal utility of income. Thus, agents who value income more highly than other agents are assigned lower values for a given health status improvement and are discriminated against in the allocation of treatments.

Healthy Year Equivalents

The healthy year equivalent is a function of the marginal rate of substitution between morbidity and longevity, much as the willingness-to-pay measure is a function of the marginal rate of substitution between health and income. This is supposed to ensure that individuals with the same preference ordering over any two health states are treated equally with respect to these two health states. However, the HYE is

constructed such that the marginal rate of substitution is evaluated at perfect health, and this can destroy the egalitarian nature of the statistic if agents differ in their risk aversion to health.

Lemma 2: The HYE value for any given health status improvement is less the more risk averse with respect to health is the agent.

Proof: see Appendix E.

Equal entitlement is not assured if agents differ in their risk aversion to health and the HYE statistic is used to allocate treatments. The more risk averse an agent is, the lower the value assigned to his or her health status improvement, and the less likely he or she will be assigned a beneficial treatment. Thus, agents who are very risk averse to health outcomes are discriminated against.

3.5 Choices Between Programs

In this section, attention is turned to the use of health status statistics to evaluate broadly based health care programs. These choices involve comparing different treatments which affect different people (e.g. pancreatic transplants affect the health of severe diabetics, lung transplants affect the health of people with severe respiratory diseases). These assessments involve measuring the health of (sometimes large) groups of people. Whether these measurements are consistent with social preferences is now assessed.

Assumptions

To make the analysis more tractable, a number of assumptions are made. First, the number of people in society is fixed (at N) so that issues of optimal population size need not be addressed. Each person is characterized by a health state

(q_i, T_i) and income level (I_i) . The social profiles of these characteristics are then $Q = (q_1, \dots, q_N)$, $T = (T_1, \dots, T_N)$, and $I = (I_1, \dots, I_N)$. It is assumed all individuals face constant prices of purchased goods, p , but may earn person-specific wage rates, w_i .

The purpose here is to determine the welfare properties of the aggregate statistics as currently used, not to invent new statistics that satisfy certain welfare properties. Because N is assumed fixed, all three aggregate health status measures are the sum of the individual statistics over all members of the community:

- a) HK $HK(Q, T) = \sum_i HK_i(q_i, T_i)$
- b) WTP $cv(Q, T) = \sum_i cv_i(q_i, T_i)$
- c) HYE $HYE(Q, T) = \sum_i HYE_i(q_i, T_i)/N$

The policy maker chooses the program which results in the highest valued aggregate health statistic. For instance, if there were twenty people willing to pay twenty thousand dollars each for pancreatic transplants, and ten people willing to pay thirty thousand dollars each for lung transplants, the aggregate value of a pancreatic transplant program would be worth one hundred thousand dollars more than a lung transplant program by the willingness-to-pay measure and would be selected by the policy maker.

Evaluation Criteria

The choices made on the basis of these aggregate health status measures should be consistent with social preferences (so that programs which society prefers most are the first to be adopted). This requires that the health status indexes themselves be consistent with such preferences.

Social preferences order health status profiles of the community. Such an ordering must possess basic consistency properties: completeness, transitivity, and asymmetry.

The evaluation criteria restrict the characterization of these social preferences. Such preferences must be welfarist (depend only on individual assessments of well-being) and sensitive to the prevailing aversion to social inequality in the community. It is assumed such social preferences can be represented by a social welfare function (SWF). This SWF must be an ethically flexible Bergson-Samuelson SWF to satisfy the above restrictions.

The aggregate health status statistics are assessed to determine (i) if they are supported by any set of rational social preferences, (ii) whether these social preferences are welfarist, and (iii) the ethical position of these preferences.

3.5.1 Rationality of the Social Ordering

The first step is to determine if the social ordering of states implied by these functions is complete, transitive, and asymmetric. The ordering is complete if the statistic is increasing in each individual's utility (i.e. is Pareto inclusive) and, therefore, in every aspect of health which contributes to utility (an incomplete ordering is called a quasi-ordering). An ordering is transitive if, when state A is weakly preferred to state B, and B to C, then A is weakly preferred to C. An ordering is asymmetric if, when A is preferred to B, B is never preferred to A.

Human Capital

The social preferences which support the human capital health status statistic (recall that the aggregate human capital statistic used in health policy and examined here is the sum of the individual statistics: $HK(Q, T) = \sum_i HK_i(q_i, T_i)$) are typically incomplete, excluding those health effects which do not affect earned income and those people not in the formal labour market (either because of institutional constraints or co-morbid conditions). Furthermore, in the unconstrained case, the index is exact

only if q is fixed. Hence, inter-morbidity comparisons cannot be made. Thus, the human capital statistic is supported by only a quasi-ordering.

Willingness-to-Pay

Recall that the aggregate willingness-to-pay statistic evaluated here is the one that is typically used in health policy: the sum of the individual willingnesses-to-pay. Willingness-to-pay statistics are complete over all health states and can be chosen to be Pareto inclusive (if the summation is defined over all individuals in society). They are not, in general, supported by social orderings because of serious rationality problems of asymmetry (see Boadway [1974], Blackorby and Donaldson [1985]), because the statistics are conditioned on end state variable values (hence, the function comparing the move from the before to the after state may differ from the function used to compare the move from the after to the before state, so that both moves may appear to be welfare enhancing in aggregate). Obviously, this is not an issue if the index is independent of these factors. Such independence is also necessary for the ordering to be welfarist, the subject of the next section, and the conditions for independence are covered there.

Healthy Year Equivalents

The aggregate healthy year equivalent health status statistic is supported by a quasi-ordering, being transitive and reflexive (the reference points are fixed for all comparisons), but not necessarily complete (while Pareto inclusive, the index is undefined for states worse than death). The severity of incompleteness is less than in the human capital case, and probably occurs infrequently in practice.

3.5.2 Welfarism

A social welfare function is welfarist if it depends only on individual assessments of well-being. This requires not only that each individual statistic be exact, but that the aggregate statistic depends only on the utility information in the individual statistics, not any conditioning or reference factors (since these factors are supposed to be irrelevant to the rankings of social states).

Roberts (1980) has derived conditions for the sum of compensating variations to represent a Bergson-Samuelson SWF. He does this in two steps. First, he shows that the ordering implied by the aggregate statistic is consistent with a Bergson-Samuelson SWF if and only if it is independent of the reference variables (prices). Second, if the social decision statistic is additive, then the social preferences which support it are independent of reference prices if and only if the indirect utility function is of the form $V^i(p, I_i) \doteq a(p)I_i + b^i(p)$. These results are now generalized to the utility based statistics considered here.

Proposition: the social ordering which is implied by a given health statistic is independent of reference variables if and only if the ordering is consistent with a Bergson-Samuelson SWF.

Proof: see Roberts (1980) or Chapter 4, Theorem 1.

Proposition: given that the aggregate health status statistic is the sum of the individual statistics, the social preferences which support the aggregate statistic are independent of reference variables if and only if

$$HS_i(u_i, x_i, y) = a(y)\gamma_i(u_i) + b_i(x_i, y), \quad (3.20)$$

where x_i are variables that may be person-specific and y are variables that are the same for all individuals.

Proof: see Roberts (1980) or Blackorby and Donaldson (1985) for the original proof based on compensating variations.

The proof is easily modified for the health index case. Consider an arbitrary index, HS_i , defined over utility, u_i , and some reference variables, θ . Then independence requires that $\sum_i HS_i(u_i^A, \theta) \geq \sum_i HS_i(u_i^B, \theta)$ for all θ . Set $\theta = \bar{\theta}$ and define $HS_i(u_i, \bar{\theta}) = \gamma_i(u_i) = z_i$, where γ_i is increasing and z_i is continuous. Then the independence condition may be restated: $\sum_i HS_i(u_i, \theta) = \sum_i HS_i(\gamma_i^{-1}(z_i), \theta) = \sum_i h_i(z_i, \theta) = H(\sum_i z_i, \theta)$, the solution to which is a Pexider equation of the form $HS_i(u_i, \theta) = a(\theta)\gamma_i(u_i) + b_i(\theta)$ (obviously, no person-specific elements of θ may appear in $a(\cdot)$).

The implications of these propositions for the three health statistics under consideration are now examined.

Human Capital

In the case of human capital, there are two cases where the statistic may be exact: when only longevity changes (and labour supply is increasing in longevity) in the unconstrained case, and when the only effect is on the labour constraint in the labour constrained case.

Corollary 1: The unconstrained human capital statistic is consistent with (3.20) if and only if (1) wages are fixed and the same for all individuals and (2) if there is no utility in leisure (e.g. $L^* = T$). The labour constrained human capital statistic is consistent with (3.20) under the same conditions as the unconstrained statistic.

Proof: see Appendix E.

In both cases, independence requires that wages are the same for all individuals and that no individual gets any utility from leisure (i.e. spends as much time as possible working). Obviously, these conditions do not hold in the real world, so the social preferences which support the human capital statistic cannot be welfarist.

Willingness-to-Pay

The willingness-to-pay statistic is the one Roberts (1980) originally dealt with. One important difference between Roberts' work and the problem addressed here is that health status (q_i, T_i) is person-specific (different people have different health).

Corollary 2: The willingness-to-pay statistic is consistent with (3.20) if and only if income and health are additively separable in the indirect utility function and utility is homothetic in income, i.e.

$$V^i(q_i, T_i, I_i, y, x_i) \stackrel{o}{=} a(y)I_i + b_i(q_i, T_i, y, x_i).$$

Proof: see Appendix E.

Unlike the Blackorby-Donaldson (1985) result for person-specific prices (that no preferences exist which satisfy independence when prices are person-specific and the aggregator function is linear), preferences do exist that satisfy independence when health status is person-specific. This is because there are no *a priori* restrictions on the relationship between health and income in the indirect utility function as there are between income and prices. Even so, empirical evidence suggests the preferences required for independence are not observed in practice (Viscusi and Evans [1990]).

Healthy Years Equivalents

In the case of healthy year equivalents, the fact that health may be person-specific is offset by the fact that the index is based on a common and fixed reference point.

Corollary 3: The healthy year equivalent statistic is consistent with (3.20) if and only if income and health are additively separable in the indirect utility function and utility is homothetic in length of life, i.e.

$$V^i(q_i, T_i, I_i, y, x_i) \stackrel{o}{=} a(q_i, y)T_i + b_i(I_i, y, x_i).$$

Proof: see Appendix E.

This restriction implies that the marginal utility of time in any given morbid state must be constant and the same for all individuals (since any morbid state may be the reference health state). This is in contrast to the willingness-to-pay case where the marginal utility of income is restricted to be *constant* and equal for all individuals. Empirical observation suggests neither set of restrictions holds in practice (in the healthy year equivalent case, discounting over time must be ruled out).

3.5.3 Ethics

The final consideration is the ethical position implied by these health status statistics. The degree of inequality aversion is reflected in the amount of curvature in the aggregator function. Each of the aggregate health status indexes is additive, reflecting inequality neutrality over the health status measure. This does not reflect inequality neutrality in the supporting social preferences, which are defined over individual utilities, not health status measures.

Human Capital

Consistency with welfarist social preferences imposes restrictions on individual preferences such that $L^* = T$. This in turn implies $V(p, w, T, I) \stackrel{\circ}{=} (wT + I)a(p)$ (or $V(p, w, \tau_L, I) \stackrel{\circ}{=} (w\tau_L + I)a(p)$). Hence, $\sum_i HK_i$ is cardinally related to the sum of individual incomes, and the human capital statistic is indifferent to inequality in income.

Willingness-to-Pay

Blackorby and Donaldson (1985) have shown that the aggregate willingness-to-pay measure used in health policy (the sum) is indifferent to income inequality. The gist of their argument (which is supported by a formal proof) is that the willingness-to-pay measures changes in well-being and therefore cannot be sensitive to inequality aversion which requires comparing levels of well-being across people.

Healthy Year Equivalents

The sum of the individual healthy year equivalents is a (non-strictly) concave function of a convex function of well-being. This is because preferences over time are assumed to be concave so that the transformation of utility imposed by the healthy year equivalent is convex. But a concave function of a convex function is not generally concave, so the inequality aversion (neutrality in this case) built into the aggregator function will not represent the inequality aversion of the social preferences which support this health statistic. In fact, the social preferences which support this index are very likely to be characterized by equality affection, in which case the health statistic assigns higher values to health profiles that are less equal.

The only exception to this situation is when individual preferences are all homothetic in length of life. Then the healthy year equivalent is a linear transform of utility and the sum of the individual healthy year equivalents is supported by social preferences that are indifferent to the distribution of well-being (see Chapter 4, Theorem 6 for a formal discussion).

3.5.4 Summary

None of these health status statistics satisfactorily orders all possible community health profiles. The human capital statistic cannot evaluate programs that affect quality of life. It could not, for instance, compare a kidney transplant program (which generally improves the quality of life but does not increase the quantity of life of kidney patients on dialysis) against a heart transplant program (which primarily increases the length of life of heart patients) in a way consistent with individual preferences. The healthy year equivalent is also incomplete and cannot assess programs that may result in health states worse than death (e.g. bone marrow transplants that reject). But the most problematic of the health status measures is the willingness-to-pay, which does not consistently rank projects (i.e. it may indicate both that a heart transplant program yields more health than a kidney transplant program, and that a kidney transplant program yields more health than a heart transplant program).

Generally, independence from extraneous factors is not achieved by any of the measures. While the theoretical requirements for independence are different for the three statistics (one could argue that they are most stringent for the human capital measure and least for the healthy year equivalent), empirical evidence indicates no set of restrictions is satisfied. Thus, no measure is supported by welfarist social preferences and the rankings of programs depend on the choice of reference variables.

Ethically, the human capital and willingness-to-pay measures are indifferent to inequality. The healthy year equivalent is the most perverse statistic, however, since its measure of health status is a convex transformation of well-being associated with health. Thus, health improvements to people who are already relatively healthy are assigned greater values than health improvements to people who are relatively unhealthy.

3.6 Conclusion

QALY analysis was developed to provide policy makers with a decision statistic that was consistent with individual preferences, but free of the ethical ramifications of linking value to ability to pay. The results of this paper suggest the success of this endeavour has been somewhat limited. The healthy year equivalent is exact, and therefore appropriately indicates which treatments are most beneficial to a given individual. In choosing individuals for treatment, observed preference patterns indicate the healthy year equivalent may actually discriminate against people who have poor health endowments (the reverse of the willingness-to-pay statistic), and may not be any less discriminatory against the financially poor than the willingness-to-pay statistic. Furthermore, it discriminates against people who are risk averse with respect to health. In evaluating health care programs, the healthy year equivalent is supported by rational social preferences, an improvement over the other measures. The consequences of the fact that these preferences are not welfarist is minimized by (in practice) the use of a fixed reference health state for all comparisons. However, the ethics of the aggregate measure may be perverse if individual utility functions are concave in time and may actually favour very unequal distributions of health over more equal distributions. Overall, the healthy year equivalent does seem to be a superior measure of health status over the alternatives available, although it is not without its own difficulties.

Additional work needs to be done before these results may be considered conclusive. Two avenues of empirical work need to be undertaken. First, the nature of preferences must be empirically identified, particularly how the marginal utilities of time and income vary with income, age, and co-morbid effects. Second, the three statistics should be assessed for relative variance and any systematic bias that arises

in their calculation in practice, and the theoretical results should be reconsidered given this information.

Theoretically, this assessment could be strengthened in a number of ways. Decisions of this nature are inherently dynamic and involve variable population sizes, and the present analysis has ignored these facts. The few comments made in this chapter about the effect of dynamic contexts on the results, and the fact that many of the results hinge on the relationship between morbidity and longevity in preferences, suggest this is one avenue deserving of further investigation. The effects of externalities should also be considered. Finally, the analysis should be extended to include the other dimensions of these rationing decisions, most notably costs.

Chapter 4

A QALY Based Societal Health Statistic for Canada, 1985

4.1 Introduction

4.1.1 Purpose

This paper assesses the appropriateness and feasibility of QALYs (quality adjusted life years) as a foundation for an index of societal health. Just as G.N.P. acts as a measure of the economy's performance in the aggregate, such an index of societal health could act as a gauge of the performance of the health care system as a whole. Such an index is necessary to assess the overall allocation of health care resources, and to effectively target policy to improve both the level and distribution of health status in society.

Results in this paper suggest that, theoretically, the QALY serves as an imperfect measure of societal health, but that these imperfections are far less severe than those associated with currently used indexes, and that such failures are probably endemic to any index based on individual preferences. A QALY based index, constructed using the best available data, indicates that morbidity has a significant effect on Canadian health status (e.g. Canadians, as a group, are prepared to give up 10 per cent of their longevity in order to eliminate morbidity from their society – more if they are inequality averse), that the distribution of health across regions and gender shifts when morbidity is accounted for (e.g. the advantage women have over men in terms of life expectancy falls by half when morbidity is taken into account, the people of

Quebec are relatively much healthier once morbidity is included in the inter-regional assessment, but people in the Atlantic provinces are worse off), and that resources might be best targeted on alleviating role (the ability to fulfil social functions) and motor (broadly defined mobility) dysfunction.

4.1.2 Criteria for a Societal Health Index

To be useful, a societal health index must assign higher values to more preferred community health status profiles than to less preferred profiles. To do this, an index must satisfy the following conditions.

Completeness

An index must encompass all aspects of health status and the health of all members of society. Otherwise, improvements (deteriorations) in the excluded aspects or in the health of the excluded individuals are not reflected in higher (lower) values of the index, even though relevant changes have occurred.

Completeness presupposes a clear definition of what health status is. Well-being in health may be distinguished from the broader concept of utility by limiting the range of aspects to those that the health care system attempts to (as opposed to does) impact directly (Evans [1984], p. 5) – assuming all other aspects of well-being can be held fixed. Such a definition encompasses those interpretations of illness relating to physical and psychological pathology, as well as broader dysfunction. More importantly, it is necessary if the index is to be used to judge the performance of the health care system.

Health status evolves over the lifespan. Point- and period-of-time indexes cannot incorporate the longevity aspect of health. Since many health care resources are

expended to reduce mortality, only lifetime health status measures are complete.¹

A societal health index must encompass all members of society and be increasing in each individual's health. This requires that the index be Pareto inclusive, though not necessarily egalitarian (where the amount of the increase is constant for all individuals), as has sometimes been suggested.

As a corollary, note that completeness also requires that the index be sensitive to relevant changes within the domain. Otherwise, the index may not detect important health status changes that have occurred.

Consistency

An index must be consistent with the set of value judgements pertaining to health status held by a benevolent social planner: health status changes which are trivial in their impact must not be assigned greater weight than those that are more important to society. The question of whose values should count and by how much is essentially normative, involving a value judgement itself.

The first normative judgement made here is that the individual who endures the health state is the best judge of the well-being associated with that state. Paternalism, on the other hand, can lead to the situation where the societal index indicates an improvement in health status even when every individual feels worse off than he or she was before.² This is an explicitly welfarist position which is well established in economics and has received increasing support in the medical literature (see, for instance, Geigle and Jones [1990]).

¹Notice that a purely outcome approach, which excludes aspects of process (how outcomes are achieved) and prognosis, has been adopted. This reflects the *ex post* position, where only health outcomes achieved matter, rather than the *ex ante* position, where all possible future health states, not just those actually attained, are included in the assessment.

²Note that the individual values the health status *outcome*, so that agency relationships, which arise from complicated processes, need not apply.

These individual valuations must then be aggregated to a societal level. This must be done in a manner consistent with social preferences (Patrick [1976]). Otherwise, the health status index may indicate that it is better to devote all resources to effect a very small health status improvement in a less deserving person (in “society’s” opinion) than a very large improvement in the health status of many more deserving individuals. Wagstaff (1991) has suggested the aggregate QALY index be treated as a social welfare function. Yet, he gives no consideration to the assumptions inherent in this approach, or the resultant social ethics. It is therefore necessary to consider what social ethics may be compromised by this approach and the implications of these compromises.

Ethical Content

The social welfare function chosen must be sensitive to ethics regarding justice and fairness held by the social planner. In terms of outcomes³, such ethics can usually be couched in terms of the distribution of health status. An index must be increasing in each individual’s health because the health of each individual is socially valuable. Also, there appears to be a trend in Canadian society towards preference for health status distributions that are more equal rather than less equal (see the evolution of this principle from the Royal Commission on Health Services [1964], where only health maximization mattered, through to the Epp Report [1986], where equality was the first of three policy priorities).

³Notice that a function defined over outcomes can only take an ethical position over outcomes. If justice is based on something else, such as the fairness of the process by which health outcomes are determined, such an index cannot reflect the ethics involved. This suggests the validity of the social welfare functional approach may have to be reconsidered at a later date.

Feasibility

Finally, an index must be feasible to construct. This condition really acts as a constraint on the optimization problem defined by the above conditions. An index cannot be considered worthwhile if the resources required for its construction are greater than any resource savings that could ensue from the use of the index as an instrument to better allocate resources. An index that does not fit the criteria above may still direct the health care system to better resource allocation than the index which satisfies the criteria but is impractical to apply. This criterion may change over time as more data and better techniques in measurement are developed.

4.1.3 Literature Review

The practice of health status measurement has followed closely contemporary data availability. The first index to be widely used is life expectancy. Its most serious drawback is incompleteness: it ignores non-fatal illness. It is consistent with preferences over longevity only if there are no states worse than death (because the index is strictly increasing in time). It is Pareto inclusive, covering all members of society, and the ethics are inequality neutral since the index is invariant to the distribution of years of life across members of society. Data are available from census information and are highly reliable.

The second class of indexes incorporate labour force data with life expectancy. These include Sullivan's (1966) index (using occupational disability), Chiang's (1968) index (using worker absenteeism), and Miller's Q (1970) (using wages lost due to illness). While such measures do account for some aspects of morbidity, they are less complete in other dimensions since only the health of persons in the labour market is counted. Furthermore, health effects which do not have an effect on worker

productivity are still ignored. The values assigned to states are inconsistent with most preferences because states in which a person is unable to work are assigned the same value as the state where that individual is dead, regardless of the utility associated with other aspects of life, whereas ability to work is deemed equivalent to perfect health, regardless of the discomfiture associated with the morbid state. Ethically, these measures are seriously flawed since the worthiness of a person's health is tied to his/her labour force productivity. Data are readily available but may be biased due to incentives on the part of workers to misreport.

The third class of indexes attempts to adjust life expectancy figures with morbidity values (such as Patrick et al.'s Q.W.B. (quality of well-being index) [1973], and Torrance's QALY [1976c]) to construct a quality adjusted life expectancy index that incorporates both length and quality of life. These measures encompass an even broader range of morbid effects, although completeness is limited to the scope of the measure in question (the Q.W.B. scale has a fixed set of components; the QALY is generic, since it can be based on any set of components). Valuations are consistent with preferences over morbidity only if the value weights are obtained from a representative sample of the population; none of the measures accommodates preference variation within the population. In order for the morbidity weighted life expectancy values to be consistent with the value of health status, the morbidity values must be appropriately weighted in terms of the value of being alive. Only the QALY explicitly does this. Individual scores are then aggregated by some function (usually additive), with the cardinality of the index dictating the nature of the comparability across utilities (e.g. when the indexes are added, this imposes the social valuation that one year of life is worth the same regardless of to whom it accrues). Data for these indexes are not readily available: even though value weights for states have been derived, the assessment of the prevalence of these states in the population has only recently been

attempted.

There have been two attempts to employ the data available to approximate an index of this class. Erickson et al. (1989) cross-reference the morbid states surveyed in the National Health Interview Survey (N.H.I.S.) to the Q.W.B. scale to construct a morbidity adjustment factor which is then combined with life expectancy figures for the U.S. They encounter a number of difficulties including: (1) The information collected in the N.H.I.S. does not correlate perfectly with the components of the Q.W.B. scale. Since the Q.W.B. scale is inclusive (measures must exist for all components and only those components), this requires proxies for some states not in the N.H.I.S. and the exclusion of other states described in the N.H.I.S. (2) One has to assume the preferences underlying the Q.W.B. scale are representative of the sample in the N.H.I.S. (3) No linkage exists between the morbidity data and the life expectancy data (i.e. the transition probabilities for moving from one morbid state to another or from any morbid state to death are unknown). Erickson et al. assume independence throughout the transition matrix (describing the probabilities of moving between morbid states over time)⁴ and match mean values from the two data sets. Furthermore, no linkage exists between the values associated with the morbid states and death. This cannot be overcome by statistical assumptions or by better data collection.

Wilkins and Adams (1983) use the data in the Canada Health Survey (1981) and purport to link these to QALY values. The statistical problems affecting Erickson et al. are still present, since the data collection is similar. The difference between the two is how these states are valued, Wilkins and Adams adopting a system that potentially is theoretically (values are linked to life) and practically (no states are

⁴They have to assume that the likelihood of occupying a given health state in any period is the same for every individual regardless of the health status of that individual in the previous period.

excluded) better. However, the morbid states chosen are very crude, resembling the labour force models of the previous decade in scope. The valuations, while they may be in a reasonable range, are arbitrary since preferences for the states used have not been surveyed. Many of the conclusions are sensitive to these choices.

While it appears the QALY based index, in principle, may be superior to other indexes, it has yet to be implemented in a fashion that achieves its advantages. The remainder of this paper assesses the theoretical performance of the QALY based index against the criteria stated above, examines the feasibility of constructing such an index with available data and what improvements in data collection may be required if current standards are inadequate, and what sorts of policy decisions may be assisted by such information to judge if these improvements are warranted.

4.2 Model

4.2.1 Theoretical Assessment

The environment in which the health states exist is described by the following assumptions.

i) An individual's state is described by x . x includes health status (q describing the morbidity profile over life, t the length of this life) and non-health factors (denoted by κ , including such aspects as income and personal characteristics). Thus, $x = (q, t, \kappa)$.

ii) It is assumed each individual ($i = 1, \dots, N$) has a well-defined preference ordering, \mathcal{R}^i , over states of the world and that this ordering has a representation, $\tilde{U}^i : R^m \rightarrow R^1$, such that

$$\tilde{U}^i(x^1) \geq \tilde{U}^i(x^0) \leftrightarrow x^1 \mathcal{R}^i x^0. \quad (4.1)$$

It is assumed this function may be conditioned on non-health factors such that $U^i(q, t)$

represents preferences over (q, t) conditioned on κ (if preferences over (q, t) are separable from κ , then this function is independent of κ , not conditioned on these variables). The value of the health state is represented by u_i ($u_i = U^i(q, t)$). It is assumed U^i is continuous in t . It is further assumed that, when necessary, these individual utility functions are interpersonally comparable and numerically measurable.

iii) It is assumed social preferences exist over the distribution of health states across the N individuals in society. It is further assumed that these social preferences are based on individual utilities. Let \mathcal{R} denote these social preferences and assume they may be represented by $W : R^N \rightarrow R^1$. Let $Q = (q_1, \dots, q_N)$ and $T = (t_1, \dots, t_N)$. Then

$$\begin{aligned} (Q^A, T^A) \mathcal{R} (Q^B, T^B) &\leftrightarrow W(u_1^A, \dots, u_N^A) \geq W(u_1^B, \dots, u_N^B) \\ &\leftrightarrow W(U^1(q_1^A, t_1^A), \dots, U^N(q_N^A, t_N^A)) \geq W(U^1(q_1^B, t_1^B), \dots, U^N(q_N^B, t_N^B)) \\ &\leftrightarrow \tilde{W}(q_1^A, \dots, q_N^A, t_1^A, \dots, t_N^A) \geq \tilde{W}(q_1^B, \dots, q_N^B, t_1^B, \dots, t_N^B). \end{aligned} \quad (4.2)$$

A social welfare function is said to be welfarist if it depends only on individual preferences or utilities alone, and is extra-welfarist if it depends on factors in addition to (u_1, \dots, u_N) . The former is often described as a Bergson-Samuelson social welfare function (B-S SWF).

iv) Assume the QALY, $\varphi(q)$ is based on the time trade-off instrument, i.e.

$$\varphi^i = \frac{m_i}{t_i} \ni U^i(q_i, t_i) = U^i(\bar{q}, m_i) \quad (4.3)$$

(where \bar{q} is the same for all individuals and is usually perfect health). Assume the time frame used in the time trade-off exercise is equal to time alive such that

$$\varphi^i t_i = m_i = M^i(q_i, t_i, \bar{q}), \quad (4.4)$$

which is the quality adjusted life time (QALT), or healthy year equivalent (Mehrez and Gafni [1989]), defined over the whole life-cycle. Note that M^i is ordinally equivalent

to U^i . Thus,

$$m_i = M^i(q_i, t_i, \bar{q}) = f_i(U^i(q_i, t_i), \bar{q}). \quad (4.5)$$

Such assumptions ensure the QALY based societal health status index is not distorted by imperfections in the valuation of individual health status alone, but only by flaws inherent in the aggregate index.

Note that $U^i(q_i, 0)$ is, logically, invariant to the level of q_i (because the individual does not endure the state, but dies immediately). Hence, because the individual is indifferent to zero years in perfect health and zero years in any other health state, he or she is not prepared to give up any time to move between these two morbidity levels. Thus, $M^i(q_i, 0, \bar{q}) = 0$. This is described as

Condition N: $M^i(q_i, 0, \bar{q}) = 0$ for all q_i .

v) The societal health index may be expressed by some aggregator function, Γ , over the QALT's:

$$HS = \Gamma(m_1, \dots, m_N). \quad (4.6)$$

Γ is supported by a set of preferences⁵ over health states, denoted by \mathcal{R}_Γ , such that

$$(Q^A, T^A) \mathcal{R}_\Gamma (Q^B, T^B) \leftrightarrow \Gamma(m_1^A, \dots, m_N^A) \geq \Gamma(m_1^B, \dots, m_N^B). \quad (4.7)$$

Whether these preferences are consistent with those of a Bergson-Samuelson social welfare function is now addressed.

The properties of Γ , given M^i, U^i , and W , are now assessed.

4.2.2 Completeness

Incompleteness over health states may occur if states worse than death exist and the QALY is bounded from below by zero. In this case, the QALY value is undefined,

⁵Torrance (1986) suggests Γ should be additive (i.e. $\Gamma = \sum_{i=1}^N m_i$), claiming such a measure is egalitarian. The accuracy of this claim is addressed later in this paper.

since no non-negative value of m solves the time trade-off problem. Whether such a situation is apt to hold over an entire lifetime is doubtful (especially if suicide is an option), so the problem may be more academic than practical.

To overcome this problem, the following regularity condition is imposed:

Condition C: $U^i(q_i, t_i)$ is increasing in t_i for all q_i .

This requires that the domain of q be restricted such that all health states are preferred to death.

The domain of the aggregator function is chosen to encompass all members of society. Because time trade-off values exist only for those members of society who are able to express their preferences, or for whom appropriate proxy values exist, there is a risk that the very young or the severely incapacitated may be misrepresented.

4.2.3 Consistency

By assumption (iv) above, the QALT measure is consistent with each individual's preferences over morbidity and mortality as conditioned on non-health factors (see Lemma 1 of Chapter 2). Thus, only consistency with social preferences need be assessed.

Since the QALTs measure levels rather than changes in health, any aggregation of QALTs represents some transitive and reflexive ordering over health status profiles over society (Γ cannot be a function of social preferences if intransitivities occur). The ordering need not be consistent with the social values of the benevolent social planner.

Dependence on Reference Health States

Recall that

$$m_i = f_i(U^i(q_i, t_i), \bar{q}) = M^i(q_i, t_i, \bar{q}), \quad (4.8)$$

so that

$$\begin{aligned}\Gamma(m_1, \dots, m_N) &= \Gamma(f_1(U^1(q_1, t_1), \bar{q}), \dots, f_N(U^N(q_N, t_N), \bar{q})) \\ &= \Gamma(M^1(q_1, t_1, \bar{q}), \dots, M^N(q_N, t_N, \bar{q}))\end{aligned}\quad (4.9)$$

and

$$(Q^A, T^A) \mathcal{R}_\Gamma (Q^B, T^B) \leftrightarrow \Gamma(M^1(q_1^A, t_1^A, \bar{q}), \dots, M^N(q_N^A, t_N^A, \bar{q})) \geq \Gamma(M^1(q_1^B, t_1^B, \bar{q}), \dots, M^N(q_N^B, t_N^B, \bar{q})). \quad (4.10)$$

This implies the social ordering, \mathcal{R}_Γ , depends on the choice of the reference health state used in the time trade-off exercise.⁶

This is clearly an undesirable property: while one health status profile might be preferred to another under one reference point, the ordering may be reversed under another reference point. Even though the actual states and satisfaction levels achieved do not change, the ordering of the two states is reversed. While the reference point is usually fixed at perfect health, so that such asymmetries do not occur in practice, it is perverse that the ordering of states should depend on one state which is not among those ranked. It implies that the aggregator function cannot be purely welfarist, so that health states cannot be evaluated simply by the well-being associated with them.

Roberts (1980) addresses a similar problem evaluating income distributions with reference prices. He found independence requires individual and social preferences to interact in a particular way. The problem may be restated for health states: Independence requires that the ordering of health status profiles be invariant to choice of reference point,

$$\Gamma(f_1(U^1(q_1^A, t_1^A), \bar{q}), \dots, f_N(U^N(q_N^A, t_N^A), \bar{q})) \geq$$

⁶It is this reference point which makes the individual valuations cardinal and allows their aggregation (in Arrow's model, the independence of irrelevant alternatives is violated). This imposed comparability invokes a certain ethical position between individuals (this is discussed in the section on ethics).

$$\begin{aligned}
& \Gamma(f_1(U^1(q_1^B, t_1^B), \bar{q}), \dots, f_N(U^N(q_N^B, t_N^B), \bar{q})) \\
& \leftrightarrow \Gamma(f_1(U^1(q_1^A, t_1^A), \bar{q}), \dots, f_N(U^N(q_N^A, t_N^A), \bar{q})) \geq \\
& \Gamma(f_1(U^1(q_1^B, t_1^B), \tilde{q}), \dots, f_N(U^N(q_N^B, t_N^B), \tilde{q}))
\end{aligned} \tag{4.11}$$

for all \bar{q}, \tilde{q} in the set of available reference points.

This condition restricts the nature of \mathcal{R}_Γ , the social preferences described by the decision statistic, Γ . These restrictions are summarized in the following five theorems.

Theorem 1: The ordering, \mathcal{R}_Γ , is independent of \bar{q} (the reference health state) if and only if it is consistent with a Bergson-Samuelson social welfare function.

Proof: see Appendix F.

In addition, independence in the social preferences places joint restrictions on the functional form of the decision statistic and individual preferences. Consider the following theorems, which describe what individual preferences must be, given four popular ethical positions on distribution.

Theorem 2: Given that Γ is additively separable, i.e.

$$\Gamma(m_1, \dots, m_N) \stackrel{o}{=} \sum_i \phi^i(m_i) \tag{4.12}$$

(where “ $\stackrel{o}{=}$ ” means “is ordinally equivalent to” such that the two expressions are related by an increasing monotonic transform), where each ϕ^i is continuous and increasing, then \mathcal{R}_Γ is independent of the reference point, \bar{q} , if and only if

$$U^i(q_i, t_i) \stackrel{o}{=} a(q_i)\phi^i(t_i) + b_i(q_i) \tag{4.13}$$

for all $i = 1, \dots, N$.

Proof: see Appendix F.

Theorem 3: Given that Γ is linear (inequality neutral) and symmetric, i.e.

$$\Gamma(m_1, \dots, m_N) \stackrel{o}{=} \sum_i m_i, \tag{4.14}$$

and that Condition N holds, \mathcal{R}_Γ is independent of the reference point, \bar{q} , if and only if

$$U^i(q_i, t_i) \stackrel{o}{=} a(q_i)t_i \quad (4.15)$$

for all $i = 1, \dots, N$.

Proof: see Appendix F.

Theorem 4: Given that Γ is symmetric Cobb-Douglas, i.e.

$$\Gamma(m_1, \dots, m_N) \stackrel{o}{=} \prod_i^N m_i, \quad (4.16)$$

\mathcal{R}_Γ is independent of the reference point, \bar{q} , if and only if

$$U^i(q_i, t_i) \stackrel{o}{=} b_i(q_i)t_i^{a(q_i)} \quad (4.17)$$

for all $i = 1, \dots, N$.

Proof: see Appendix F.

Theorem 5: Given that Γ is maximin (extreme inequality aversion), i.e.

$$\Gamma(m_1, \dots, m_N) \stackrel{o}{=} \min\{m_1, \dots, m_N\}, \quad (4.18)$$

and that for each \bar{q} , the range of $M^i(\cdot, \cdot, \bar{q})$ is the same for all i , then \mathcal{R}_Γ is independent of the reference point, \bar{q} , if and only if

$$U^i(q_i, t_i) \stackrel{o}{=} U^k(q_k, t_k) \quad (4.19)$$

for all $i = 1, \dots, N$. That is, all individuals have identical preferences.

Proof: see Appendix F.

Notice the restrictions on individual preferences implied by the above. The value of an extra year in the reference health state must be the same, regardless of to whom it accrues. Unless the choice of reference state can be limited, this restriction must

apply to all health states.⁷ This assumes a degree of preference similarity across the individuals in society.

Normally, the preferences of individuals must be taken as given and the social preferences are chosen. If, as is widely believed, time is discounted geometrically, the only ethical position above that can be taken without inducing reference point dependence is the Rawlsian, and this only if preferences are identical.

These results are disappointing since they suggest, given realistic assumptions about individual preferences, that there is no aggregator of QALTs that is consistent with a Bergson-Samuelson social welfare function.⁸

4.2.4 Ethical Content

The previous section suggests that the ethical flexibility of the social welfare function is limited by the structure of individual preferences *if* welfarism is maintained. If the welfarism condition is relaxed, can the ethical position of a QALY based societal health status index be made more defensible?

At a minimum, the social welfare function should be increasing in each individual's level of health. But the ethical position should also incorporate some amount of inequality aversion: a society in which some individuals live long healthy lives while

⁷In Chapter 2, the extended sympathy QALY instrument was developed. It differed from the established QALY instruments in its use of a reference individual as well as a reference health state. Hence, resultant QALY comparisons are consistent with interpersonal utility comparisons. In this case, conditions for independence apply only to the reference individual's utility function: in the additive SWF case, independence requires $U^r(q, t) = \phi^r(a(q)t + b(q))$ (i.e. the reference individual's utility function must be quasi-homothetic). While this condition must apply to all individuals, since any individual may be chosen to be the reference individual, no cross-person restrictions apply (e.g. it is not necessary that $a^i = a^j$). Hence, the conditions for independence when the extended sympathy instrument is used are less restrictive than when any of the established instruments are used. Only the extended sympathy case can reflect that the marginal value of a year of life in any given health state can be of different value to different people.

⁸This is quite apart from the concern of Wagstaff (1991) that such a function is defined over only a partial welfare space.

others live short and miserable lives is less preferred to one where all individuals endure a moderate amount of good health. To reflect such a position, the aggregator function must be S-concave⁹ across levels of individual health status.

Inequality Aversion

The first issue is whether the index can incorporate inequality aversion. If it cannot, the index is inadequate for any policy evaluations where the distribution of outcomes is considered important.

The aggregator function can easily be chosen so as to be concave in its arguments (the QALTs). This implies inequality aversion to the distribution of well-being only if the QALT is concave in time (since a concave function of a concave function is itself concave, but a concave function of a convex function need not be). This again depends on the structure of preferences.

Theorem 6 (Blackorby and Donaldson [1988]): $M^i(q_i, t_i, \bar{q})$ is concave in t if and only if

$$U^i(q_i, t_i) \stackrel{o}{=} a_i(q_i)t_i + b_i(q_i). \quad (4.20)$$

Proof: see Appendix F.

Since preferences are probably strictly concave in t (as they are under when time is discounted geometrically), the aggregation of QALTs may imply perverse social preferences where social states characterized by greater inequality are assigned higher values than states with less inequality.

Combining the results from Theorems 2 to 5 and 6, the only ethical position which satisfies inequality aversion and welfarist principles is the utilitarian (i.e. $\Gamma = \sum M^i(U^i, \bar{q})$), and this only under the assumption that individuals do not discount

⁹If the social welfare function, W , is symmetric and quasi-concave, then it satisfies S-concavity. W is quasi-concave if its level sets are convex.

over time.

Horizontal Equity

Torrance (1986, p. 17) has argued in favour of the arithmetic mean as an aggregator function, claiming it is ethically just because it assigns equal weight to a year in perfect health for each individual. However, if preferences vary, the aggregation involved is horizontally inequitable.

QALY analysis imposes $\tilde{M}(\bar{q}, 1, \alpha^i) = M^i(\bar{q}, 1, \bar{q}) = 1$, where \bar{q} is perfect health and α^i describes the personal characteristics on which preferences are conditioned. In addition, the individual selects a morbid state deemed equivalent to death (a variable reference point) such that $\tilde{M}(q(\alpha^i), 1, \alpha^i) = \tilde{M}(q, 0, \alpha^i) = 0$, where $q(\alpha^i)$ is the death equivalent. Suppose there are two agents, A and B, and that A is more averse to death than B (i.e. $q(\alpha^A) < q(\alpha^B)$). Assuming both individuals' preferences are homothetic in time, so that the conditions for Theorem 3 are satisfied and the aggregator function reflects the inequality neutral equity position,

$$m_i = \frac{a^i(q) - a^i(q(\alpha^i))}{a^i(\bar{q}) - a^i(q(\alpha^i))}. \quad (4.21)$$

Assume only one element of α^i , ξ , affects this choice of death equivalent and differentiate with respect to this element:

$$\text{sign} \frac{\partial \tilde{M}^i}{\partial \xi^i} = -\text{sign} \frac{\partial q(\alpha^i)}{\partial \xi^i} \quad (4.22)$$

and

$$\text{sign} \frac{\partial^2 \tilde{M}^i}{\partial q \partial \xi^i} = \text{sign} \frac{\partial q(\alpha^i)}{\partial \xi^i}. \quad (4.23)$$

This reflects that agent A is more likely to prefer projects that increase longevity, while B is more likely to prefer projects which improve the quality of life. Suppose A and B occupy the same health state and that a project which improves quality of

life may be given to only one of them. Both would prefer to have the project, since, for a given length of life, both prefer better morbid states to worse morbid states. Yet B will always be chosen to receive the project even though the aspect over which preferences vary (death) is unaffected by the project. Even though linearity exists in individual and social preferences, egalitarianism (defined as equal entitlement to the same resources by people who occupy the same observed state) is not ensured. This is an example of how ethics may be capricious when cardinality is assumed to impose comparability.

4.2.5 Summary

It is clear that aggregated QALTs do not generate completely satisfactory indexes of health status, even when individual QALTs are measured in a way completely consistent with individual preferences. Identical individual preferences are neither necessary nor sufficient for aggregated QALTs to reflect acceptable social preferences. Of greatest concern is that the curvature in individual preferences may cause aggregated QALTs to favour extreme distributions of health status, rather than more equal distributions. This contradicts the egalitarian spirit with which such QALY based indexes were first developed (Torrance [1986]).

Problems of incompleteness and cross-person assessments which are inconsistent with interpersonal utility comparisons can be circumvented by slight modifications in the measurement instrument; but the other concerns raised above are inherent in the QALY approach. Thus, some ethical compromises must be made. However, such compromises are characteristic of all well-being based indexes, and those associated with QALY based health indexes are more agreeable than most of the alternatives. Whether it is feasible to implement an index of this nature is now assessed.

4.3 Empirical Assessment

This section examines the feasibility of constructing a QALY based societal health index given current data availability. The theoretical model described above requires a tremendous amount of data since it is based on the value of the morbidity path over the entire lifetime. Such data are not available, nor are they apt to be in the near future. This section describes what approximations must be accepted if only currently available data are used in the construction of such an index.

4.3.1 Data

Data available include life expectancy (Statistics Canada [1991]), a point in time morbidity profile, stratified by age (Statistics Canada [1987]), and preferences over these morbid states. These data represent all adult (age 15 and over) Canadians living in the ten provinces (the territories are excluded). These data are supplemented with data on persons living in institutions (Statistics Canada [1990]). Morbidity aspects include endurance, agility, and perception (long-term physical ill-health), role and socio-emotional function, and short-term ill-health. The path of morbidity over the lifetime remains as yet unknown. Data sources and the following procedure are described in greater detail in the data appendix (Appendix G).

4.3.2 Procedure

Briefly, the procedure is as follows

- i) QALY weights for the non-institutionalized population are calculated from the General Social Survey data: First, reports of health satisfaction are compared to reports of morbidity factors, including chronic ill-health (endurance, perception, and agility), short-term ill-health, social ill-health and emotional ill-health. Each factor

is represented by a binary structure and may be either present or absent and, if present, severe or not severe. The construction of these morbidity factors is discussed in appendix G. Observations with multiple chronic or short-term morbid states are deleted so that estimates reflect the marginal disutility of the morbid state as taken from a point of perfect health and not any cross-effects from other illnesses. The function is estimated by Probit analysis because the satisfaction responses are reported as ordered categories (it must be assumed that all individuals use the same time frame and the same cardinal scale for these responses). These estimates are given in Table G.1. Satisfaction values for all possible configurations of morbid states are then reconstructed using a functional form chosen to be consistent with a multiplicative multi-attribute utility function.¹⁰ Second, morbid states in the G.S.S. are then matched with QALY values reported in the literature. Ten such matches were found (these are given in Appendix G). These QALY values were then regressed against the above estimated satisfaction values associated with these states, with the restriction that perfect health be assigned a value of one imposed. A logarithmic functional form was found to provide the best fit of this estimated relationship. All satisfaction values derived in the first step are transformed according to this estimated relationship so that they are consistent with a time trade-off scale. This procedure is repeated separately for men and women since their estimated satisfaction functions are found to be significantly different from each other. These transformation functions are given in Appendix G as well.

ii) Because the sampling methods used in the G.S.S. excluded persons in institutions, including those institutionalized because of ill-health, the G.S.S. data provide an over-estimate of the health of Canadians. To adjust for this bias, numbers of persons institutionalized for ill-health by age, province, and sex are taken from the

¹⁰The discussion in Chapter 2 explains why this estimation procedure is appropriate.

H.A.L.S. (1990) and are incorporated into the morbidity data base (these figures are adjusted for population growth between 1985 and 1986 and for persons living in the territories to make the population base comparable with that of the G.S.S. – see Appendix G for details). The state of being institutionalized is then assigned a range of values found in the literature (.33 to .56).

iii) Measures of morbidity for five and ten year age groupings are then calculated. Morbid states endured by people falling within these age groups are weighted by their estimated QALY values and summed. The arithmetic means of these QALY-weighted morbidity values are calculated using the population weights in the G.S.S. and the adjusted population counts from the H.A.L.S.

iv) The expectation of living to a given age is calculated, conditioned on being alive at age fifteen (this is necessary to make the population base comparable with that of the G.S.S.). These expectations are then summed according to the same age groupings on which the average estimated QALYs are based.

v) Life expectancy is then weighted by the average QALY value for each age grouping, i.e.

$$HSE = \sum_{t=15}^{80} \left(\frac{1}{2} P^D(t|15) + P^S(t|15) \right) \left(\sum_{i=1}^{N_t} \hat{\varphi}(q_{i|t}) / N_t \right), \quad (4.24)$$

where HSE is the estimated health status index, $P^D(t|15)$ is the probability of dying in the t th year given the individual lived to age 15 (this is multiplied by $1/2$ on the assumption that people, on average, die at the mid-point of the time interval), $P^S(t|15)$ is the probability of surviving to the t th year given the individual survived till age 15, N_t is the number of people in age category t for whom morbidity data are available (weighted as described above), and $\hat{\varphi}(q)$ is the estimated QALY value for health state q . Standard errors for these estimates are approximated using a δ -method as described in Rothenberg (1984), under the assumption of independence

between the estimators of the value of health states (the estimated QALYs) and the prevalence of those states (the arithmetic averages).

HSE is an estimate of the quality adjusted life expectancy for adult Canadians living in the provinces. How good an estimate depends on the appropriateness of the following three preference assumptions, necessary to statistically link the available data sets to construct a societal health index.

i) Additivity of Preferences.

The value of the lifetime path for morbidity must be constructed by adding the values of health states endured for subperiods of life, rather than valuing the lifetime health status as a whole. The bias associated with this procedure depends on the structure of preferences.

Lemma 1: The sum of QALYs defined over periods of life less than full life is equivalent to the QALT if and only if preferences may be represented as

$$U(q, t, \kappa) = \phi(a(q, \kappa)t). \quad (4.25)$$

Proof: see Appendix F.

Since observed preferences are typically concave in time, the sum of QALYs typically over-estimates the true quality adjusted lifetime.

ii) Independence from Time Frame

G.S.S. satisfaction levels are regressed on QALY values in the literature. The latter are based on a time frame of 70 years. The former are not dimensioned by a fixed time. In fact, there is no reason to believe the time frames used by respondents do not vary. If responses depend on time, such a procedure is biased.

Lemma 2: QALYs are independent of time if and only if preferences may be represented by

$$U(q, t, \kappa) = \phi(a(q, \kappa)t). \quad (4.26)$$

Proof: see Appendix F.

Since QALYs are believed to be concave in time, QALY values estimated over longer time frames will be less than QALYs estimated over shorter time frames. Since the G.S.S. satisfaction responses are probably based on some time frame less than the lifespan, the estimated function which transforms the G.S.S. preferences to QALY values will produce QALY estimates that are biased downwards.

iii) Strict Independence Must Exist Between (1) Mortality and Morbidity in Social Preferences or (2) in Their Bivariate Joint Distribution.

Since the true bivariate distribution is unknown (the G.S.S. data are not linked to the mortality data), one either has to combine time alive and morbidity in a fashion that is independent of this distribution or make assumptions about this distribution given the aggregator function chosen.

Consider first the restrictions which must be imposed on the functional form of the SWF if no restrictions are imposed on the bivariate distribution. Given that the true SWF, W , is defined over QALTs and the estimated SWF, \hat{W} , is defined, in the absence of data linking the occurrence of morbidity with mortality, over a QALY index, $I(\varphi(q))$, and a life expectancy index, $J(t)$, then unbiasedness ensues if and only if

$$W(\varphi(q_1)t_1, \dots, \varphi(q_N)t_N) = \phi(\hat{W}(I(\varphi(q)), J(t))). \quad (4.27)$$

But this is a Pexider equation with the solution

$$W(\cdot) = a \prod (\varphi(q_i)t_i)^{b_i}. \quad (4.28)$$

Let b_i be the proportion of people in states $(\varphi(q), t)$, $\varphi(q)$, and t respectively. Then the condition for unbiasedness may be expressed

$$\prod \varphi(q_i)t_i^{Pr(\varphi(q)t)} = \prod (\varphi(q_i))^{Pr(\varphi(q))} \prod t_i^{Pr(t)}. \quad (4.29)$$

Taking logarithms yields

$$\sum_t \sum_q Pr(\varphi(q), t) \ln(\varphi(q_i) t_i) = \sum_q Pr(\varphi(q)) \ln \varphi(q_i) + \sum_t Pr(t) \ln t_i. \quad (4.30)$$

Using the properties of logarithms and probabilities, (4.30) may be re-expressed as

$$\begin{aligned} \sum_t Pr(t) \ln(t_i) + \sum_q Pr(\varphi(q)) \ln(\varphi(q_i)) = \\ \sum_q Pr(\varphi(q)) \ln \varphi(q_i) + \sum_t Pr(t) \ln(t_i). \end{aligned} \quad (4.31)$$

This result holds regardless of the joint distribution between morbidity and mortality. However, the solution is based on lifetime QALYs (which are unavailable). If the function is based on QALT segments, each segment is treated as a different but far less well off person and the Cobb-Douglas social preferences underestimate the true value of societal health. Thus, this solution is not very practical.

Alternatively, one can begin with a particular social welfare function and determine what sorts of conditions must be imposed on the distribution function. Since the inequality neutral SWF is invariant to the use of piecemeal QALY values, begin with the function

$$W(\varphi(q)_1, \dots, \varphi(q)_N, t_1, \dots, t_N) = \int_t \int_{\varphi(q)} (\varphi(q)t) f(\varphi(q), t) d\varphi(q) dt = E(\varphi(q)t), \quad (4.32)$$

where $f(\varphi(q), t)$ is the joint distribution function. The estimated function is

$$\bar{W} = \left(\int_{\varphi(q)} \varphi(q) f(\varphi(q)) d\varphi(q) \right) \left(\int_t t f(t) dt \right) = E(\varphi(q)) E(t). \quad (4.33)$$

But $E(\varphi(q)t) = E(\varphi(q))E(t) \leftrightarrow Cov(\varphi(q), t) = 0$ (i.e. the distribution of morbidity and mortality are independent).

If morbidity and mortality are positively correlated (i.e. well people live longer), then the estimated health status index overestimates the true value of health status in society.

With these caveats in mind, attention is now turned to the calculated quality adjusted life expectancies.

Table 4.1: Quality Adjusted Life Expectancy, Canada and the Provinces

men	CDA	NWFD	PEI	NS	NB	QUE
$E(t)$	58.98	58.85	58.39	58.22	58.44	57.88
$E(\varphi t)$	53.94 (4.45)	52.88 (7.36)	52.00 (8.56)	52.38 (6.97)	53.26 (6.87)	53.84 (7.42)
	ONT	MAN	SASK	ALTA	BC	
$E(t)$	59.32	59.07	59.90	59.57	60.03	
$E(\varphi t)$	53.74 (7.76)	53.61 (6.86)	55.08 (7.13)	53.87 (6.64)	55.40 (7.22)	
women	CDA	NWFD	PEI	NS	NB	QUE
$E(t)$	65.52	65.13	65.93	64.89	65.80	65.14
$E(\varphi t)$	57.70 (4.39)	56.19 (7.19)	58.67 (8.61)	56.45 (6.86)	57.36 (7.45)	57.79 (7.29)
	ONT	MAN	SASK	ALTA	BC	
$E(t)$	65.47	65.73	66.47	65.82	66.13	
$E(\varphi t)$	57.34 (7.23)	57.01 (6.74)	57.56 (7.52)	58.24 (7.17)	59.30 (7.40)	

4.4 Results and Implications

4.4.1 Quality Adjusted Life Expectancy

The calculated quality adjusted life expectancies are given in Table 4.1 (standard errors are in brackets where applicable). Life expectancy in Canada at age fifteen is 58.98 years for men and 65.52 years for women (ranging, for men, from 57.88 years in Quebec to 60.03 years in British Columbia and, for women, from 64.89 years in Nova Scotia to 66.47 years in Saskatchewan). After adjusting for quality, the national figures become 53.94 years for men and 57.70 years for women (ranging, for men, from 52.00 years in P.E.I. to 55.40 years in British Columbia and, for women, from 56.19 years in Newfoundland to 59.30 years in British Columbia).

These figures suggest that morbidity is an important component of ill health and that the population as a whole would be willing to give up ten per cent of life

expectancy to eradicate ill-health while alive.¹¹ It is interesting to note the differential impact of the adjustment across the provinces. Before adjusting for quality, Quebec has the poorest average level of health, while the westernmost provinces have the best health indicators. After adjusting for quality, Quebec ranks fourth among the ten provinces, while, among men, only British Columbia maintains a clear advantage in the west, and the Atlantic provinces prove to be at a greater health risk than is indicated by mortality alone.^{12 13}

Although the average falls after quality adjustment, the variance increases (the range nearly doubles). This may reflect different policies towards institutionalizing the severely disabled across the provinces. Since this effectively withdraws the sicker members from the sample used in the survey, the above estimates may be biased. To check this, the adjustment is repeated with observations on the disabled in institutions accounted for. These results are presented in Table 4.2¹⁴ (again, standard errors appear in brackets where applicable).

The effect of the institutional adjustment on rankings is minimal. However, the adjustments are directly related to the healthiness of the population (i.e. the higher

¹¹It is possible to make such claims because morbidity has been measured using a time trade-off instrument. Hence, the final index indicates how much time is worth how much morbidity because morbidity is measured in terms of time. This is one of the principle advantages of using QALY data rather than another health status index in the calculation of a societal health index.

¹²Such differences are not due to variations in tolerance for certain states (since average preferences are used) but actual advantages in morbidity.

¹³One must be cautious when interpreting these figures since the standard errors of these estimates, particularly for the smaller provinces, tend to be quite high. The national figures are based on a much larger sample and are correspondingly that much more reliable. For this reason, most of the ensuing analysis focuses on national figures. It is interesting to note that the greater part of these high standard errors is driven by high variance in the morbid states achieved and not the estimated values of these states. This reinforces the argument that the arithmetic mean is an unsatisfactory index of societal health because it is incapable of reflecting this wide dispersion of achieved outcomes, and that additional data must be collected to allow such distributionally sensitive measurements.

¹⁴The unadjusted figures may differ between Tables 4.1 and 4.2 because aggregation occurs over 10 year periods instead of 5 year periods. If morbidity increases at higher ages, one would expect the 10 year averages to be higher than the 5 year averages. The results indicate this holds in most cases.

Table 4.2: Quality Adjustment with Institutional Data, Canada and the Provinces

men	CDA	NWFD	PEI	NS	NB	QUE
$E(t)$	58.98	58.85	58.40	58.22	58.44	57.88
$E(\varphi t/\text{no I})$	53.94 (4.45)	52.64 (7.36)	52.71 (8.56)	52.42 (6.97)	53.32 (6.87)	53.82 (7.42)
$E(\varphi t/\varphi(I) = .56)$	53.72 (4.38)	52.46 (7.57)	52.53 (8.55)	52.27 (6.98)	53.14 (6.85)	53.58 (7.48)
$E(\varphi t/\varphi(I) = .33)$	53.52 (4.38)	52.37 (7.57)	52.41 (8.55)	52.15 (6.98)	52.99 (6.85)	53.37 (7.48)
	ONT	MAN	SASK	ALTA	BC	
$E(t)$	59.32	59.07	59.90	59.57	60.03	
$E(\varphi t/\text{no I})$	53.72 (7.76)	53.78 (6.86)	55.15 (7.13)	53.91 (6.64)	55.45 (7.22)	
$E(\varphi t/\varphi(I) = .56)$	53.50 (7.71)	53.63 (7.04)	54.94 (7.02)	53.65 (6.56)	55.17 (7.14)	
$E(\varphi t/\varphi(I) = .33)$	53.30 (7.71)	53.49 (7.04)	54.75 (7.02)	53.39 (6.56)	54.95 (7.14)	
women	CDA	NWFD	PEI	NS	NB	QUE
$E(t)$	65.52	65.13	65.95	64.89	65.80	65.14
$E(\varphi t/\text{no I})$	57.75 (4.39)	56.47 (7.19)	58.34 (8.61)	56.52 (6.86)	57.56 (7.45)	57.92 (7.29)
$E(\varphi t/\varphi(I) = .56)$	57.40 (4.32)	56.25 (7.29)	58.13 (8.64)	56.29 (6.74)	57.10 (7.41)	57.52 (7.33)
$E(\varphi t/\varphi(I) = .33)$	56.97 (4.32)	55.95 (7.29)	57.82 (8.64)	55.98 (6.74)	56.66 (7.41)	57.10 (7.33)
	ONT	MAN	SASK	ALTA	BC	
$E(t)$	65.47	65.73	66.47	65.82	66.13	
$E(\varphi t/\text{no I})$	57.34 (7.23)	57.15 (6.74)	57.58 (7.52)	58.28 (7.17)	59.33 (7.40)	
$E(\varphi t/\varphi(I) = .56)$	57.04 (7.11)	56.82 (6.88)	57.30 (7.51)	57.84 (7.04)	58.94 (7.30)	
$E(\varphi t/\varphi(I) = .33)$	56.60 (7.11)	56.43 (6.88)	56.91 (7.51)	57.35 (7.04)	58.52 (7.30)	

the quality adjusted health status of the province, the higher the rate of institutionalization), with the two most western provinces and Quebec showing the highest proclivity to institutionalize, and the Atlantic provinces showing the least. Thus, while differential rates of institutionalization account for part of the differential, they are not sufficient to explain away the patterns observed. Furthermore, the adjustment is greater for women than for men in every case, suggesting that women are institutionalized at a much higher rate and that the non-institutionalized quality adjusted figures are biased in favour of the women.

4.4.2 Male-Female Differentials

The above results suggest that, while women may live longer than men, they may not live as well. In fact, the advantage is nearly halved between the quality adjusted figures and the unadjusted figures (the unadjusted life expectancies give women a 6.54 year advantage over men; this falls to 3.76 years after adjusting for non-institutional morbidity and 3.45 years after adjusting for institutionalization as well).

A residual test indicates preferences for health states differ significantly between men and women (see Appendix G for details: women generally associate more disutility with short term ill-health and severe cases of ill-health than men, while men associate greater disutility with chronic and more moderate cases of ill-health). For this reason, the estimation procedure is repeated for each group separately. The linkage to the QALY values is complicated by the fact that the QALY data are averaged over both sexes (supposedly because the values do not differ between the two groups - see Torrance et al. [1982] and Torrance [1976b]). Estimated satisfaction levels are converted to estimated QALY values both by the common transformation used before and by sex-specific transforms that are estimated on the assumption that QALY values are the same for the two groups (the latter of these two methods is probably

Table 4.3: Male-Female Differentials, Canada

	no adjust.	jt. pref. jt. trans.	spec. pref. jt. trans.	spec. pref. spec. trans.
men	58.98	53.94 (4.45)	53.12 (4.85)	54.01 (5.74)
women	65.52	57.70 (4.39)	57.05 (5.23)	55.73 (4.34)
difference	6.54	3.76	3.93	1.72

the more theoretically sound – see the data appendix for details). These results are presented in Table 4.3 (standard errors appear in brackets).

Using common preference functions and common QALY transforms, the non-institutional difference is 3.76 years. With specific preference functions and common transforms, the difference is 3.93 years. With specific preference functions and specific transforms, the difference falls to only 1.72 years – a quarter of the unadjusted differential. One possible interpretation of these results is that women are prepared to give up about twice as much longevity as men in order to spend their remaining years in perfect health. This is in part due to the fact that women experience more illness while alive than men (the difference is apparent when common values are used to weight morbid states), and in part because women seem to place greater value on living well rather than longer (the differences are greater when gender-specific values of morbid states are used).

While these results must be accepted cautiously because of the high standard errors, they do suggest that, overall, more resources should be devoted to women's health than had heretofore been considered appropriate. Especially in medical research directed at alleviating ill-health rather than prolonging life, women should be given more consideration than men, not less.

Table 4.4: Morbidity by Major Category, Men

illness	prevalence present	prevalence severe	disutility present	disutility severe	societal health all
endurance	14.22	2.56	.1300 (.006)	.0087 (.115)	1.14 (.176)
role	10.40	1.19	.1818 (.013)	.3175 (.095)	1.49 (.385)
emotional	3.20	N/A	.2206 (.021)	N/A	.500 (.215)
social	12.71	N/A	.0183 (.003)	N/A	
hearing	8.55	.72	.0071 (.006)	.1041 (.153)	.280 (.259)
sight	3.37		.0626 (.023)		
short	8.02	5.15	.0551 (.009)	.1184 (.011)	.810 (.247)
agility	5.82	.80	.0643 (.023)	.1386 (.651)	.25 (.406)

4.4.3 The Importance of Morbidity

It appears that morbidity significantly detracts from health in Canada. This is hardly surprising for an industrialized country. But if more resources are to be targeted at the alleviation of morbidity, where should they best be directed? The impact of morbidity has been assessed in terms of prevalence and the disutility of any given factor, but not in terms of social disutility. Results for all three measures are given in Table 4.4 and 4.5. (Note: prevalence is the percentage of the population with the condition, disutility is the estimated marginal amount of dissatisfaction taken at perfect health, and societal health is the mean number of quality adjusted life years that is achieved if the condition is eradicated.)

Note that the social and perception values for societal health include both social and emotional, and hearing and sight components respectively.

Table 4.5: Morbidity by Major Category, Women

illness	prevalence present	prevalence severe	disutility present	disutility severe	societal health all
endurance	23.67	2.56	.0725 (.003)	.1471 (.042)	2.17 (.180)
role	12.47	1.19	.1412 (.013)	.1801 (.057)	2.10 (.388)
emotional	3.56	N/A	.1468 (.017)	N/A	.560 (.203)
social	9.59	N/A	.0070 (.003)	N/A	
hearing	7.00	.92	.0721 (.010)	.0965 (.153)	.330 (.217)
sight	4.58		.0273 (.018)		
short	12.90	7.12	.0928 (.005)	.1535 (.007)	1.33 (.246)
agility	9.26	.86	.0011 (.017)	.3772 (.254)	.40 (.380)

Problems of endurance are the most prevalent, affecting nearly 20 per cent of the Canadian population, while endurance and short-term ill-health are the most prevalent severe illnesses. Social ill-health is relatively more prevalent among men, while short term ill-health is relatively more prevalent among women. Emotional ill-health is the least prevalent morbid effect for both groups.

The estimated disutility is greatest for morbidity in the role category, and smallest in the perception category. For men, the role category is far more important than for women. Social and emotional function are also more important. Women, on the other hand, are far more concerned about agility and relatively more concerned with short-term episodes of ill-health.

The aggregate effects are, as to be expected, a combination of the two results above. The most significant category for men is role, while for women, it is endurance

(although both categories rank highly for both men and women). The least important categories are agility for men and perception for women (again, both categories are relatively unimportant for both men and women). The societal index gives a clear indication of where resources should be targeted to alleviate morbidity. While the components of the endurance category fall plainly in the medical realm, one has to consider if the role category could be better approached through more socially oriented programs. This lends support to the idea that health and welfare programs should be integrated to achieve the greatest returns.

4.5 Avenues for Future Research

This paper has identified a number of problems associated with QALY based measures of societal health. The supposed advantage of such measures is their relationship to preferences, both individual and social. Yet these relationships have been shown to be flawed.

The ranking of different health status profiles may depend on some health state not even under consideration. The comparability assumptions imposed by this reference state are also unsatisfactory. More distressing, the QALY based index may be inconsistent with the distributional ethics held by the benevolent social planner – even when the aggregator function is chosen to accommodate these ethics. The inconsistencies between the QALY based index and social preferences are theoretical in nature and cannot be overcome by improved data collection. One should recognize that social orderings of this nature are invariably flawed and that the flaws associated with this index have been identified and may well be the best compromise available.

QALY based measures may also distort the preferences of individuals, although this problem is empirical rather than theoretical, and can be overcome by improved

data collection. The first problem is to identify the paths of morbidity over lifetimes. While it is impractical to wait for every member of society to complete his or her morbidity path, longitudinal surveys (such as the Canada Sickness Survey [1954]), rather than the periodic ones currently planned, could be used to estimate the transition matrix over time and the joint distribution over morbidity and mortality. The second problem is to collect valuation data over these states. The exercise carried out above is not completely satisfactory since satisfaction levels are reported on an arbitrary scale that has no linkage to the value of life and may vary across individual respondents. Furthermore, values constructed from periodic QALYs (rather than QALTs) are apt to be biased and correction factors need to be estimated. This requires a much broader set of QALY values than currently exists. Finally, if the collected QALY data support the hypothesis that QALY values differ across people, or that different individuals value health differently (or should have their health valued differently from the rest of the population), the assumption of identical preferences must be relaxed and an index with interpersonal comparability properties must be adopted instead of the time trade-off.

Chapter 5

Conclusion

One of the problems in health care evaluation is the measurement of health outcomes. “Natural” or clinical units (such as number of cases of a disease in a population) are unsatisfactory because they have no underlying value basis. It is impossible to compare different health conditions or different changes in the same condition. Standard economic measures (such as willingness-to-pay) are, on the other hand, badly distorted by imperfections in health and related markets (e.g. supplier induced demand), and favour allocations that are biased to the wealthy. Such values may not reflect the true “worth” of a health state, and the worthiness of an individual’s health depends on his or her income.

The QALY (quality adjusted life year) has been put forth as a solution to this measurement problem. It is a health index whose weights are based on preferences for health states (hence, it has a value foundation), and these values are free of many of the less attractive features of the standard economic measures. QALYs are now used to (a) choose treatments for a given individual (e.g. cancer therapy), (b) choose individuals for a given treatment (e.g. organ transplants), and (c) choose programs for funding (e.g. Oregon Medicaid reforms, Ontario formulary lists). Unfortunately, it is not clear if such allocations are appropriate. While there has been considerable research done on the intrapersonal properties of QALYs (i.e. whether they are utility numbers or not), there has been an inadequate investigation of the interpersonal and aggregation properties that are involved, particularly in the latter two types of

decisions. This thesis investigates these neglected welfare properties.

The first essay addresses two issues of how to obtain QALY values: how to value health states and how to identify which health states must be valued. The first goal of this chapter is to identify valid measurement instruments (which convert preferences for health states to a numerical scale). The definition of construct validity employed explicitly incorporates how QALY values are used to make policy decisions. Two necessary conditions for validity are derived from this concept: QALY values must represent the individual's preferences over health states and QALY values should be independent of the level of non-health factors. All known QALY instruments are then evaluated for consistency with these conditions. All QALYs are found to rank health states appropriately, but none can rank improvements (or changes) in health (without imposing unrealistic restrictions on preferences). Nor does any QALY instrument generate values that are independent of context. It is concluded that there is no "gold standard" instrument. It is shown that the presumed theoretical correctness of the standard gamble instrument (or the theoretical incorrectness of the time trade-off instrument) is wrong. Policy makers must instead choose the QALY instrument which fits the type of decision to be made. For decisions involving different lengths of life, a time trade-off instrument should be used; while for decisions involving different numbers of people, a person equivalent instrument should be used. In general, the description of the reference state should be expanded to include a fixed level for every aspect of the state that could change as a result of the health care projects being compared.

One significant contribution made in this chapter is the development of the extended sympathy instrument. The literature to date has focused on the measurability properties of QALY values, and comparability properties, which are involved whenever choices between individuals are made, have been neglected. Unlike any other

instrument, the extended sympathy is based on comparable utility functions. The values obtained are thus appropriate whenever the policy decision involves choosing between individuals. The extended sympathy instrument also provides one method to improve the interpersonal ethics involved in some QALY statistics discussed in subsequent chapters.

The second part of this chapter examines whether any of the reconstruction methods proposed by Keeney and Raiffa (1976) apply over broadly defined categories of ill-health. This is the only known example where mutual preference independence over widely applicable health conditions is treated as the null hypothesis. Additive structures are found to be biased, but multiplicative structures are found to be valid. This result suggests there can be significant cost savings in QALY data collection if a “utility index” approach is adopted. When a large variety of health conditions must be valued (for instance, when system-wide evaluations are to be undertaken, such as was done recently in Oregon), the most cost-effective method of obtaining QALY values is to identify the attributes of each condition and the value for each attribute. The value for any condition can be found just from the values for its component attributes. This result also suggests that other health status indexes, most of which have an additive structure, do not appropriately value health states.

The third chapter assesses if the types of allocation decisions suggested by QALY indexes are appropriate and whether the alternative measures (willingness-to-pay and human capital) perform any better. Different but mutually consistent ethical criteria are imposed on the different types of decisions which use these values.

When choosing a treatment for a given individual, the goal is to give the patient the health outcome he or she most prefers. Although this “patient centred ethic” is gaining widespread acceptance, it contradicts the paternalism that occasionally appears in health policy. Both the healthy year equivalent (the QALY index chosen

for this analysis) and the willingness-to-pay measure achieve this goal, but the human capital measure generally does not. Hence, if the human capital statistic is used to determine the treatment path for a patient, the patient may be left in what, in his or her opinion, is an inferior health state of those available.

When choosing an individual for a given treatment, the principle followed in this chapter is equal entitlement. Equal entitlement requires that any two individuals who have the same preferences over two health states should have an equal opportunity to achieve either health state. This is perhaps the most contentious principle used in the chapter and it is not uncommon to find other positions based on merit or health maximization. The human capital statistic is found to discriminate against the wealthy, the retired, and those who value leisure. The willingness-to-pay measure is found to discriminate against the poor, the retired, and those who value income. Finally, the healthy year equivalent is found to discriminate against the retired, those who are risk averse with respect to health, and, very likely, those who are otherwise in poor health. These results seriously undermine the egalitarian justification for using QALYs. QALYs are not only inconsistent with equal entitlement, but the people who are discriminated against are, in some cases, the very ones widely believed to deserve extra consideration (i.e. those endowed with poor health and those who have taken care of their health endowments).

One method proposed in the literature (Wagstaff [1991]) to overcome this problem is to assign “named weights” to individuals so that more deserving people carry greater weight in social decisions than less deserving people. But this method is inadequate to prioritize individuals by their health status since, when health status changes, the weight remains fixed. Instead, the position adopted in this thesis is that such apparent discrimination can be overcome by the use of a more distributionally sensitive aggregation rule. Such a rule assigns higher values to whomever is in worse

health. This requires attention be turned from measuring health status changes to health status levels. The use of health status improvement (change) measures implies utilitarian ethics, which are sometimes inconsistent with egalitarian principles. The use of health status change measures is only defensible, if at all, for small projects that have little impact on the overall distribution of health in society.

When choosing a program for funding, society wants to pick the set of programs that gives the best level and distribution of community health. Thus, a good index should encompass all aspects of community health (that is, be complete). It should measure these aspects such that community health increases whenever the health of a constituent improves. To be consistent with the first ethical principle, this requires the index be consistent with welfarist social preferences. Because, *ceteris paribus*, equal distributions of health are socially preferred to unequal distributions of health, these preferences should also be distributionally sensitive. This is consistent with the second ethical principle, which deals with horizontal rather than vertical equity. The aggregate versions of the three health statistics commonly used in practice are assessed for consistency with these objectives.

The human capital measure is found to be seriously incomplete. It is not consistent with welfarism since this requires that everyone earn the same wage and spend all their time working. It is distributionally neutral. The willingness-to-pay measure generally fails to order health profiles sensibly. It is consistent with welfarism only if the individual willingness-to-pay for a health status change is the same for all levels of income. It is also distributionally neutral. The mean (or sum of) healthy year equivalent(s) can order most health profiles sensibly (only states worse than death pose a problem). It is welfarist only if the individual healthy year equivalents are proportional to the length of life. However, the distributional ethics of this index are (likely) perverse, and the index favours unequal distributions of health in the

community. Clearly, the first two indexes are unsatisfactory, neither being able to even identify which community health profiles are better than others.

While the healthy year equivalent can order most community health profiles, the manner in which it measures them is not completely satisfactory. Typically, the index depends on the reference health state. This result simply reflects the consequences of Arrow's Impossibility Theorem: reasonable social orderings based solely on individual preferences do not exist. To combine individual orderings over health states, the aggregate QALY index converts these preferences to a cardinal scale by anchoring each individual ordering such that a year in the reference health state is the same value for everyone (i.e. the independence of irrelevant alternatives is violated).¹ It should be recognized that some such ethical compromise is involved for any index based on individual (or patient) preferences. The alternative is to adopt a paternalistic position (dictatorship) that the QALY index was designed to overcome.

The policy maker can circumvent the most serious implications of reference state dependence by ensuring that all community health measures are fixed to the same reference state (usually perfect health). Then a consistent ranking of health profiles is generated.

A more distressing ethical implication of aggregate QALY statistics is that they assign higher values to improving the health of people who are already relatively healthy rather than those who are unhealthy. To compensate for this, the policy maker can choose an aggregation rule that is characterized by a high degree of inequality aversion (the feasibility of such rules is addressed in the fourth chapter).

These results suggest that the QALY is an acceptable decision tool for choosing

¹An alternative approach to combine individual orderings is to make these orderings interpersonally comparable. The extended sympathy instrument generates such interpersonally comparable values. If the policy maker found the implications of cardinality (a year of life in perfect health is worth the same to all) too offensive, the extra investment required to collect extended sympathy QALY values would be warranted.

treatments for a given patient, is better than the alternatives for choosing between programs for funding (although the policy maker may be well advised to use an aggregation rule with strict inequality aversion to compensate for the distributional effects), but is perhaps unacceptable for choosing between people for treatment. Policy makers must be prepared to favour the healthy and those who are less cautious regarding their health if they are going to allocate resources on the basis of improvements in health measured by QALY-type statistics.

The fourth chapter of this thesis examines the theoretical and empirical properties of a QALY-based societal health index. In response to some of the issues raised in the previous chapter, different aggregation rules over individual healthy year equivalents are examined. It is investigated whether there exists any aggregation of individual healthy year equivalents that represents acceptable social preferences for community health. Aggregation is shown to be possible, but, as with any social ordering generated by individual orderings for health states, some ethical compromises must be made. It is found that the only aggregation rule which represents complete, welfarist, and distributionally defensible social preferences is the sum, but that this requires the assumption that individuals do not discount over time. Given that this condition usually does not hold in practice, some ethical compromise must be made. Most health planners would probably sacrifice welfarism to obtain more appropriate distributional ethics. However, the indexes which fall into this category are based on quality adjusted lifetimes and cannot be estimated with the piecemeal data that are available. Hence, an additive structure, with all its accompanying distributional faults, is used to estimate the health status statistic.

A measure of health status for adult Canadians living in the provinces is calculated with the available data. Results indicate (1) morbidity significantly reduces the health of Canadians (society would be willing to give up 10 percent of its life expectancy

to rid itself of all morbidity), (2) that women suffer from more morbidity than men (their health advantage is cut in half when life expectancy is adjusted for quality), and (3) that the most socially significant categories of ill-health are role and motor dysfunction. Policy makers could use information of this type to more effectively target health care resources (e.g. allocation of funds between life-saving versus life-improving interventions, allocation of funds between prevention programs for gender-specific diseases, targeting programs that assist people with moving around their community and occupational retraining programs).

This chapter demonstrates that, while the construction of such indexes is certainly feasible, it could be greatly improved with additional data. First, information on the dynamics of illness would allow a broader range of aggregation rules (and more distributionally defensible ethics) and overcome the positive bias in this index because mortality and morbidity had to be assumed uncorrelated. Second, individual health state values which are appropriately scaled to a QALY interval (rather than satisfaction with health as reported on an ordinal scale) would overcome the need to assume restrictions on individual preferences to construct the social index, and increase its empirical reliability.

This thesis has answered a number of questions relating to the use of QALYs in health policy making. Its main contribution to the literature has been to identify the underlying interpersonal and social ethics of QALY-based decision statistics that are currently used to guide health care policy. For decisions involving choices between individuals, the interpersonal ethics are shown to be non-egalitarian. In fact, the QALY index discriminates against the very people most societies would, if anything, choose to offer preferential health care access (the people born with poor endowments of health and those who have taken care of their health endowments). Thus, the interpersonal properties of the QALY index are in some ways inferior to those of the

competing health status measures they were designed to replace.

These interpersonal problems can be overcome if a decision statistic is chosen which incorporates concerns for inequality. This requires abandoning the currently used change in health measure (which, with utilitarian underpinnings, is inconsistent with egalitarianism) in favour of a social welfare function defined over individual QALYs. The “more deserving” can then be given more weight in policy decisions by choosing an aggregation rule that incorporates strict inequality aversion. The results of this thesis indicate such social aggregation is both theoretically and practically possible, although some compromise in social ethics is involved. However, these ethical compromises are inherent for any social ordering based on individual preferences (Arrow’s theorem), and must be accepted unless the policy maker is prepared to give up the position that the “patient knows best”. In order to adhere to this principle, the position taken throughout the thesis is that the individual is the best judge of his or her own welfare (this is an underlying principle in QALY analysis) and that health states should be valued by the individual enduring them (i.e. QALY values should come from patients, or from people who have the same preferences as these patients). Average community preferences should only be used to determine the nature of the aggregation rule (i.e. the degree of inequality aversion across health in the community). The use of average preferences to evaluate individual health states could result in the situation where society’s health improves by forcing patients to accept treatments (and health outcomes) that they do not want.

Current aggregation rules are shown to be inequality promoting. They also impose the social value that a year of life in perfect health (or whatever the reference health state is) is worth the same to everyone. It is shown that the first problem can be overcome by choosing an aggregation rule that is more distributionally sensitive (this is clearly possible in theory, although more data are required to implement

this). The second problem, if it is indeed considered a problem, can be overcome by using an extended sympathy QALY instrument. These values are interpersonally comparable, so it is unnecessary to impose preference similarity arbitrarily. The implied social ethics of the extended sympathy QALY-based aggregate index may or may not be more acceptable to the policy maker, and the extra resources required by this measurement technique may or may not be warranted.

This thesis has raised as many questions as it has answered. This new research agenda is summarized below. Issues of who should provide QALY values when individuals are unable or unwilling to do so are not addressed. This is not a trivial problem. Some people are too incapacitated to provide QALY values (the very young, the very old, the mentally ill), while others may recognize incentives to misreport the value of certain health states (they could under-report the value of their current health state to attract more resources to treatment of this state). One response to these problems is to use average QALY value functions (as opposed to average QALY values that entail a certain ethical position), where the average is taken from a representative group (representative in the sense that people in this group have the same preferences as the person whose health state needs to be valued). There has not been adequate empirical analysis as to whether such representative groups exist (evidence from Chapters 2 and 4 indicate that statistically significant differences do appear to exist across some characteristics), and whether responses are normally distributed so that the use of an average response is appropriate. Another line of empirical analysis that should be undertaken is to investigate (1) whether the extended sympathy instrument is feasible in practice, and (2) if so, whether responses vary significantly to justify the use of such an instrument. The discussion above should suffice as justification for this analysis. These two lines of research constitute a departure from the empirical analysis that has been undertaken up till now. The typical focus has

been on content validity (whether different instruments generate the same values). Statistical differences consistent with theory are now well established. Hence, it is argued here that field research should be taken in new directions.

The welfare assessment can be improved in a number of ways. The analysis in this thesis treated morbidity and mortality as completely separable. A more realistic representation would have morbidity evolving over the life-cycle. This complicates the analysis considerably and may change some of the results obtained in the simpler model (e.g. the healthy year equivalent function may become discontinuous). Another issue is how to incorporate other data that are relevant to the policy maker in a decision statistic, particularly resource use. The analysis in the thesis presupposed that a fixed level of funding for health care had already been determined. The issue of how, if at all, QALYs could be used to help determine what this overall budget should be was not addressed. Many of the results in the thesis suggested QALYs could be used to determine the optimal distribution (as opposed to the level) of resources across health programs by asking which generated the greatest societal health. This requires evaluating the health of the entire community, even those not affected by the project, for every decision. More feasible piecemeal rules, which approximate the best strategy plan, should be devised and assessed for accuracy.

The most exciting line of research involves extensions in the area of societal health measurement. While theoretically sound, greater confidence could be placed in the empirical results (and the policies suggested by them) with two improvements in data collection. The first is to obtain more reliable values for the health states endured by members of society. No work has been done on whether or not the health conditions recorded in the General Social Survey capture all the relevant dimensions of ill-health. It would also be worthwhile to learn how close the values assigned to the health states in this research are to the values that would have been obtained if they had been

measured directly with QALY instruments. Obviously, if these values are significantly different, the societal health measures obtained and the policy implications drawn from them may have to be adjusted. The second improvement in data collection is to find the “careers” of certain health conditions (i.e. the transitional probabilities of moving between health states over time). This would reduce the statistical bias in the calculated quality adjusted life expectancy, but would also allow the use of other aggregate statistics that reflect a greater degree of inequality aversion. Even with the data available, continued measurement of societal health can generate policy relevant information. Such measures can be repeated over time (to track the health of a population), or across regions (to compare the relative performance of different health care systems or the impact of other factors that affect the health of a community). Such data are, in fact, necessary for any type of system-wide assessment.

Bibliography

- [1] Anderson, J., J. Bush, M. Chen, and D. Dolenc (1986). "Policy Space Areas and Properties of Benefit-Cost/Utility Analysis." *JAMA*, vol. 255, 794-795.
- [2] Atkinson, A. and F. Bourguignon (1982). "The Comparison of Multi-Dimensioned Distributions of Economic Status." *Review of Economic Studies*, vol. 49, 183-202.
- [3] Arrow, K. (1978). "Extended Sympathy and the Possibility of Social Choice." *Philosophia*, 7, 223-237.
- [4] Avorn, J. (1984). "Benefit and Cost Analysis in Geriatric Care." *New England Journal of Medicine*, vol. 310, 1294-1301.
- [5] Birch, S. and C. Donaldson (1987). "Applications of Cost-Benefit Analysis to Health Care." *Journal of Health Economics*, vol. 6, 211-225.
- [6] Birch, S. and A. Gafni (1991). "Cost-Effectiveness/Utility Analyses: Do Current Decision Rules Lead Us to Where We Want to Be?" CHEPA D.P. 91-6.
- [7] Birnbaum, M. (1973). "The Devil Rides Again: Correlations as an Index of Fit." *Psychological Bulletin*, vol. 79, 239-240.
- [8] Blackorby, C. and D. Donaldson (1990). "A Review Article: The Case Against the Use of the Sum of Compensating Variations in Cost-Benefit Analysis." *Canadian Journal of Economics*, vol. 23, 471-494.
- [9] Blackorby, C. and D. Donaldson (1988). "Money Metric Utility: A Harmless Normalization?" *Journal of Economic Theory*, vol. 46, 120-129.
- [10] Blackorby, C. and D. Donaldson (1985). "Consumers' Surpluses and Consistent Cost-Benefit Tests." *Social Choice and Welfare*, vol. 1, 251-262.
- [11] Blackorby, C., D. Primont, R. Russell (1978). *Duality, Separability, and Functional Structure: Theory and Economic Applications*. New York: North-Holland.
- [12] Boadway, R. (1974). "The Welfare Foundations of Cost-Benefit Analysis." *Economic Journal*, vol. 84, 426-439.
- [13] Boadway, R. and N. Bruce (1984). *Welfare Economics*. New York: Basil Blackwell.

- [14] Bombardier, C., A. Wolfson, A. Sinclair, and A. McGeer (1982). "A Comparison of Three Preference Measurement Methodologies in the Evaluation of a Functional Health Status Index." In R. Deber and G. Thompson (eds.). *Choices in Health Care: Decision Making and Evaluation of Effectiveness*. Toronto: University of Toronto Press.
- [15] Boyle, M. and G. Torrance (1984). "Developing Multiattribute Health Indexes." *Medical Care*, 22, 1045-1057.
- [16] Boyle, M., G. Torrance, J. Sinclair, S. Horwood (1983). "Economic Evaluation of Neonatal Intensive Care of Very-low-birth-weight Infants." *New England Journal of Medicine*, vol. 308, 1330-1337.
- [17] Brent, R. (1991). "A New Approach to Valuing Life." *Journal of Public Economics*, vol. 44, 165-172.
- [18] Brooks, R. (1986). *The Development and Construction of Health Status Measures*. IHE Report 1986:4.
- [19] Broome, J. (1978). "Trying to Value a Life." *Journal of Public Economics*, vol. 9, 91-100.
- [20] Butler, J. (1990) "Welfare Economics and Cost-Utility Analysis." A.N.U. Dept. of Economics W.P. 205.
- [21] Canada. Dept. of National Health and Welfare (1954). *Canada Sickness Survey*. Ottawa: Dominion Bureau of Statistics.
- [22] Canada. Dominion Bureau (1960). *Illness and Health Care in Canada. Canadian Sickness Survey, 1950-1951*. Cat. 82-518. Ottawa: Queen's Printer.
- [23] Canada. Parliament (1964). *Royal Commission on Health Services Report*. Ottawa: Queen's Printer.
- [24] Canada. Statistics Canada (1990). *The Health and Activity Limitation Survey Highlights: Disabled Persons in Canada*. Cat. 82-620. Ottawa: Minister of Regional Industrial Expansion.
- [25] Canada. Statistics Canada (1987). *General Social Survey: Health and Social Support, 1985*. Cat. 11-612. Ottawa: Minister of Supply and Services.
- [26] Canada. Statistics Canada (1981). *Health of Canadians: Report of the Canada Health Survey*. Ottawa: Minister of Supply and Services.

- [27] Canada. Statistics Canada, Canadian Centre for Health Information (1991). *Health Reports: Life Tables: Canada and the Provinces, 1985-1987*, supp. 13, vol. 2. Ottawa: Minister of Industry, Science and Technology.
- [28] Carr-Hill, R. (1989). "Assumptions of the QALY Procedure." *Social Science and Medicine*, vol. 29, 469-477.
- [29] Carr-Hill, R. (1985). "The Evaluation of Health Care." *Social Science and Medicine*, vol. 21, 367-75.
- [30] Chew, S. (1980). *Two Representation Theorems and Their Application to Decision Theory*. Unpublished Ph.D. thesis. U.B.C.
- [31] Chiang, C. (1968). *Introduction to Stochastic Processes in Biostatistics*. New York: John Wiley.
- [32] Churchill, D., B. Lemon and G. Torrance (1984). "A Cost-Effectiveness Analysis of Continuous Ambulatory Peritoneal Dialysis and Hospital Hemodialysis." *Medical Decision Making*, vol. 4, 489-500.
- [33] Culyer, A. (1989). "The Normative Economics of Health Care Finance and Provision." *Oxford Review of Economic Policy*, vol. 5, p. 34-58.
- [34] Culyer, A. (1976). *Need and the National Health Service*. London: Martin Robertson and Company.
- [35] Deaton, A. and J. Muellbauer (1980). *Economics and Consumer Behaviour*. New York: Cambridge University Press.
- [36] Diewert, W. E. (1982). "Duality Approaches to Microeconomic Theory." In K. Arrow and M. Intriligator (eds.). *Handbook of Mathematical Economics*, vol. 2. New York: North-Holland Publishing Company.
- [37] Diewert, W.E. (1973). "Functional Forms for Profit and Transformation Functions." *Journal of Economic Theory*, vol. 6, 284-316.
- [38] Donabedian, A. (1971). "Social Responsibility for Personal Health Services: An Examination of Basic Values." *Inquiry*, vol. 8, 3-19.
- [39] Donaldson, David (1991). "On the Aggregation of Money Measures of Well-Being in Applied Welfare Economics." Paper presented to the Western Agricultural Association/American Agricultural Association.
- [40] Drummond, M. (1987). "Cost Benefit Analysis in Health Care: Future Directions." In G. Teeling Smith (ed.). *Health Economics: Prospects for the Future*. New York: Croom Helm.

- [41] Drummond, M., G. Stoddart, and G. Torrance (1987). *Methods for the Economic Evaluation of Health Care Programs*. Toronto: Oxford University Press.
- [42] Eichhorn, W. (1978). *Functional Equations in Economics*. Don Mills, Ontario: Addison-Wesley Publishing Company.
- [43] Epp, J. (1986). *A Framework for Health Promotion*. Ottawa: Health and Welfare Canada.
- [44] Erickson, P., E. Kendall, J. Anderson, and R. Kaplan (1989). "Using Composite Health Status Measures to Assess the Nation's Health." *Medical Care*, vol. 27, s66-76.
- [45] Evans, R.G. (1984). *Strained Mercy: The Economics of Canadian Health Care*. Toronto: Butterworths.
- [46] Feeny, D. and G. Torrance (1989). "Incorporating Utility Based Quality-of-Life Assessment Measures in Clinical Trials: Two Examples." *Medical Care*, vol. 27, S190-204.
- [47] Fishburn, P. (1988). *Non-linear Preference and Utility Theory*. John Hopkins University Press.
- [48] Fishburn, P. (1964). *Decision and Value Theory*. New York: Wiley.
- [49] Froberg, D. and R. Kane (1989). "Methodology for Measuring Health State Preferences - I-IV." *Journal of Clinical Epidemiology*, vol. 42, 345-354, 459-471, 585-592, 675-685.
- [50] Furlong, W., D. Feeny, G. Torrance, R. Barr, and J. Horsman (1990). "Guide to Design and Development of Health State Utility Instrumentation." McMaster University, CHEPA, D.P. 90-9.
- [51] Gafni, A. and S. Birch (1991). "Equity Considerations in Utility-Based Measures of Health Outcomes in Economic Appraisals: An Adjustment Algorithm." *Journal of Health Economics*, vol. 10, 329-342.
- [52] Gafni, A. and G. Torrance (1984). "Risk Attitude and Time Preference." *Health Management Science*, vol. 30, 440-451.
- [53] Geigle, R. and S. Jones (1990). "Outcomes Measurement: A Report from the Front." *Inquiry*, vol. 27, 7-13.
- [54] Giauque, W. and T. Peebles (1976). "Application of Multiattribute Utility Theory in Determining Optimal Test Treatment Strategies for Streptococcal Sore Throat and Rheumatic Fever." *Operations Research*, 24, 933-950.

- [55] Haldane, J. (1988). "Persons and Values." *Journal of Medical Ethics*, vol. 14, 39-41.
- [56] Hanke, S. (1981). "On the Feasibility of Benefit-Cost Analysis." *Public Policy*, vol. 29, 147-158.
- [57] Harris, J. (1987). "QALYfying the Value of Life." *Journal of Medical Ethics*, vol. 13, 117-123.
- [58] Harsanyi, J. (1955). "Cardinal Welfare, Individual Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy*, vol. 63, 309-321.
- [59] Hausman, J. and D. Wise (1978). "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica*, vol. 46, 403-426.
- [60] Hilden, J. (1985). "The Non-Existence of Interpersonal Scales: A Missing Link in Medical Decision Theory." *Medical Decision Making*, vol. 5, 215-228.
- [61] Kahneman, D. and A. Tversky (1979). "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, vol. 47, p. 263-291.
- [62] Jones-Lee, M. (1976). *The Value of Life: and Economic Analysis*. Chicago: University of Chicago Press.
- [63] Keeney, R. and H. Raiffa (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: Wiley.
- [64] Kennedy, P. (1985). *A Guide to Econometrics*, 2nd ed. Cambridge: MIT Press.
- [65] Klein, G., H. Moskowitz, S. Mahesh, and A. Ravindran (1985). "Assessment of Multi-Attributed Measurable Value and Utility Functions via Mathematical Programming." *Decision Sciences*, vol. 16, 309-324.
- [66] Klevit, H., A. Bates, T. Castanares, E. Kirk, P. Sipes-Metzler, and R. Wopat (1991). "Prioritization of Health Care Services. A Progress Report by the Oregon Health Services Commission." *Archives of Internal Medicine*, vol. 151, 912-916.
- [67] Krischer, J. (1976). "Utility Structure of a Medical Decision." *Operations Research*, vol. 24, 951-972.
- [68] Labelle, R. and J. Hurley (1991). "Implications of Basing Health Care Resource Allocation on Cost-Utility Analysis in the Presence of Externalities." CHEPA D.P. 91-3.

- [69] Lipscomb, J. (1980). "Value Preferences for Health: Meaning, Measurement, and Use in Program Evaluation." In R. Kane and R. Kane (eds.) *Values and Long-Term Care*. Toronto: Lexington.
- [70] Longmore, D. and H. Rehamn (1975). "The Cumulative Cost of Death." *Lancet*, vol. 1, 1023-1025.
- [71] Loomes, G. and L. McKenzie (1989). "The Use of QALY's in Health Care Decision Making." *Social Science and Medicine*, vol. 28, 299-308.
- [72] Luce, R. and H. Raiffa (1957). *Games and Decisions*. John Wiley and Sons.
- [73] Machina, M. (1982). "Expected Utility Analysis Without the Independence Axiom." *Econometrica*, vol. 50, p. 277-323.
- [74] Maynard, A. (1991). "Developing the Health Care Market." *Economic Journal*, vol. 101, 1277-1286.
- [75] McDowell, I. and C. Newell (1987). *Measuring Health: A Guide to Rating Scales and Questionnaires*. Oxford University Press.
- [76] McGuire, A., J. Henderson, and G. Mooney (1988). *The Economics of Health Care*. New York: Routledge and Kegan Paul.
- [77] McNeil, B. and S. Pauker (1980). "Optimizing Patient and Societal Decision Making by the Incorporation of Individual Values." In R. Kane and R. Kane (eds.). *Values and Long-Term Care*. Toronto: Lexington.
- [78] McNeil, B., S. Pauker, H. Sox, A. Tversky (1982). "On the Elicitation of Preferences for Alternative Therapies." *New England Journal of Medicine*, vol. 306, 1259-1262.
- [79] McNeil, B., R. Weichselbaum and S. Pauker (1981). "Speech and Survival Trade-offs Between Quality and Quantity of Life in Laryngeal Cancer." *New England Journal of Medicine*, vol. 305, 982-987.
- [80] Mehrez, A. and A. Gafni (1991). "The Healthy Years Equivalents: How to Measure Them Using the Standard Gamble Approach." *Medical Decision Making*, vol. 11, 140-146.
- [81] Mehrez, A. and A. Gafni (1989). "Quality Adjusted Life Years, Utility Theory, and Healthy-years Equivalents." *Medical Decision Making*, vol. 9, p. 142-149.
- [82] Miller, J. (1970). "An Indicator to Aid Management in Assigning Program Priorities." *Public Health Reports*, vol. 85, 724-731.

- [83] Mishan, E. (1976). *Cost Benefit Analysis*. New York: Praeger Publishers.
- [84] Mooney, G. (1977). *The Valuation of Life*. London: MacMillan Press.
- [85] Neu, C. (1980). "Individual Preferences for Life and Health: Misuses and Possible Uses." In R. Kane and R. Kane (eds.). *Values and Long-Term Care*. Toronto: Lexington.
- [86] Patrick, D. (1976). "Constructing Social Metrics for Health Status Indexes." *International Journal of Health Services*, vol. 6, 443-453.
- [87] Patrick, D., J. Bush, and M. Chen (1973). "Methods for Measuring Levels of Well-Being for a Health Status Index." *Health Services Research*, vol. 8, 228-245.
- [88] Patrick, D. and M. Bergner (1990). "Measurement of Health Status in the 1990's." *Annual Review of Public Health*, vol. 11, 165-183.
- [89] Pliskin, J., D. Shephard and M. Weinstein (1980). "Utility Functions for Life Years and Health Status." *Operations Research*, vol. 28, 206-224.
- [90] Read, J., R. Quinn, D. Berwick, H. Fineberg, and M. Weinstein (1984). "Preferences for Health Outcomes." *Medical Decision Making*, vol. 4, 315-329.
- [91] Roberts, K. (1980). "Price Independent Welfare Prescriptions." *Journal of Public Economics*, vol. 13, 277-297.
- [92] Rosser, R. and P. Kind (1978). "A Scale of Valuations of States of Illness. Is There a Social Consensus?" *International Journal of Epidemiology*, vol. 7, 347-358.
- [93] Rothenberg, T. (1984). "Approximating the Distributions of Econometric Estimators and Test Statistics." In Z. Griliches and M. Intriligator (eds.). *Handbook of Econometrics*, vol. 2. New York: North-Holland, 882-935.
- [94] Sackett, D. and G. Torrance (1978). "The Utility of Different Health States as Perceived by the General Public." *Journal of Chronic Diseases*, vol. 31, 697-704.
- [95] Sen, A. (1986). "Social Choice Theory." In K. Arrow and M. Intriligator (eds.). *The Handbook of Mathematical Economics*, vol. 3. New York: North-Holland.
- [96] Sen, A. (1985). *Commodities and Capabilities*. North-Holland.
- [97] Sen, A. (1972). "Control Areas and Accounting Prices: an Approach to Economic Evaluation." *Economic Journal*, vol. 82, s486-501.
- [98] Shoemaker, P. (1982). "The Expected Utility Model: its Variants, Purposes, Evidence, and Limitations." *Journal of Economic Literature*, vol. 20, 529-563.

- [99] Stevens, S. (1959). "Measurement, Psychophysics and Utility." In C. Churchman and P. Ratoosh (eds.). *Measurement Definitions and Theories*. Wiley. p. 18-63.
- [100] Sullivan, D. (1966). "Conceptual Problems in Developing an Index of Health." *Vital and Health Statistics*, vol. 2, 17.
- [101] Sutherland, H., V. Dunn, and N. Boyd (1983). "Measurement of Values for States of Health with Linear Analog Scales." *Medical Decision Making*, vol. 3, 477-487.
- [102] Torrance, G. (1987). "Utility Approach to Measuring Health Related Quality of Life." *Journal of Chronic Diseases*, vol. 40, 593-603.
- [103] Torrance, G. (1986). "Measurement of Health State Utilities for Economic Appraisal." *Journal of Health Economics*, 5, 1-30.
- [104] Torrance, G. (1982). "Multi-Attribute Utility Theory as a Method of Measuring Social Preferences for Health States in Long-Term Care." In R. Kane and R. Kane (eds.). *Values and Long-Term Care*. Lexington: Lexington Books.
- [105] Torrance, G. (1976a). "Toward a Utility Theory Foundation for Health Status Index Models." *Health Services Research*, vol. 11, 349-369.
- [106] Torrance, G. (1976b). "Social Preferences for Health States: An Empirical Evaluation of Three Measurement Techniques." *Socio-Economic and Planning Sciences*, vol. 10, 129-136.
- [107] Torrance, G. (1976c). "Health Status Index Models: A Unified Mathematical View." *Management Science*, vol. 9, 990-1001.
- [108] Torrance, G., M. Boyle, and S. Horwood (1982). "Application of Multi-attribute Utility Theory to Measure Social Preferences for Health States." *Operations Research*, vol. 30, 1053-1069.
- [109] Torrance, G., G. Stoddart, M. Drummond, and A. Gafni (1981). "Cost Benefit Analysis versus Cost-Effectiveness Analysis for the Evaluation of Long Term Care Programs." *Health Services Research*, vol. 16, 474-476.
- [110] Torrance, G., W. Thomas, and D. Sackett (1972). "A Utility Maximization Model for Evaluation of Health Care Programs." *Health Services Research*, vol. 7, 118-33.
- [111] van Praag, B. (1968). *Individual Welfare Functions and Consumer Behaviour. A Theory of Irrational Rationality*. Amsterdam: North-Holland.

- [112] Veit, C. and J. Ware (1982). "Measuring Health and Health Care Outcomes." In R. Kane and R. Kane (eds.). *Values and Long-Term Care*. Lexington: Lexington Books.
- [113] Viscusi, W. and W. Evans (1990). "Utility Functions that Depend on Health Status: Estimates and Economic Implications." *American Economic Review*, vol. 80, 353-374.
- [114] Wagstaff, A. (1991). "QALYs and the Equity Efficiency Trade-off." *Journal of Health Economics*, vol. 10, 21-41.
- [115] Ware, J. and J. Young (1979). "Issues in the Conceptualization and Measurement of Value Placed on Health." In S. Mushlin and D. Dunlop (eds.) *Health: What is it Worth?*. Toronto: Pergamon Press.
- [116] Weinstein, M. and W. Stason (1977). "Allocation of Resources to Manage Hypertension." *New England Journal of Medicine*, vol. 296, 732-739.
- [117] White, K., S. Wong, D. Whistler, and S. Haun (1990). *SHAZAM User's Reference Manual Version 6.2*. Toronto: McGraw-Hill.
- [118] Wilkins, R. and O. Adams (1983). *Healthfulness of Life*. Montreal: The Institute for Research on Public Policy.
- [119] Williams, A. (1988). "Priority Setting in Public and Private Health Care." *Journal of Health Economics*, vol. 7, 173-83.
- [120] Williams, A. (1987). "Measuring Quantity of Life." In G. Teeling Smith (ed.) *Health Economics: Prospects for the Future*. New York: Croom Helm.
- [121] Williams, A. (1983). "Economics of Coronary Artery Bypass Grafting." *British Medical Journal*, vol. 291, 326-329.
- [122] Wolfson, A., A. Sinclair, C. Bombardier, and A. McGeer (1982). "Preference Measurements for Functional Status in Stroke Patients: Interrater and Intertech-nique Comparisons." In *Values in Long-Term Care*.
- [123] Wright, S. (1985). "Health Satisfaction: A Detailed Test of the Multiple Discrepancies Theory Model." *Social Indicators Research*, vol. 17, 299-313.

Appendix A

Chapter 2 Proofs

Proof of Lemma 1 (CS):

Given the assumptions in the description section,

$$\varphi^{CS}(q; t, \kappa) = \frac{U(q, t, \kappa) - U(q^0, t, \kappa)}{U(q^1, t, \kappa) - U(q^0, t, \kappa)}. \quad (\text{A.1})$$

1.a) Ordering levels

Since q^1 and q^0 are fixed points, $U(q^1, t, \kappa)$ and $U(q^0, t, \kappa)$ are fixed values (\bar{u} and \underline{u} respectively) so

$$\varphi^{CS}(q; t, \kappa) = \frac{U(q, t, \kappa) - \underline{u}}{\bar{u} - \underline{u}} = aU(q, t, \kappa) + b \stackrel{c}{=} U(q, t, \kappa), \quad (\text{A.2})$$

where $a = \frac{1}{\bar{u} - \underline{u}}$ and $b = \frac{\underline{u}}{\bar{u} - \underline{u}}$. If $U(q^1, t, \kappa) > U(q^0, t, \kappa)$, then $a > 0$ by strict monotonicity.

1.b) Ordering differences

Since $\varphi^{CS}(q; t, \kappa) \stackrel{c}{=} U(q, t, \kappa)$, $\varphi^{CS} \in S^c(\hat{U}) \leftrightarrow U \in S^c(\hat{U})$.

2) Completeness:

Obviously, if $\bar{u} \neq \underline{u}$, then any q defined for $U(q, t, \kappa)$ is defined for $\varphi^{CS}(q; t, \kappa)$.

3) Uniqueness:

a) Invariance to $\phi(U)$. Let $U' = \phi(U)$.

Necessity: By the definition of uniqueness $\varphi^{CS}(q; t, \kappa)_U = \varphi^{CS}(q; t, \kappa)_{U'}$. Substituting for $\varphi^{CS}(q; t, \kappa)_{U'}$:

$$\varphi^{CS}(q; t, \kappa)_U = \frac{\phi(U(q, t, \kappa)) - \phi(U(q^0, t, \kappa))}{\phi(U(q^1, t, \kappa)) - \phi(U(q^0, t, \kappa))}. \quad (\text{A.3})$$

Rearranging this expression yields

$$\begin{aligned} \phi(U(q, t, \kappa)) - \phi(U(q^0, t, \kappa)) &= \\ \varphi^{CS}(q; t, \kappa)_U (\phi(U(q^1, t, \kappa)) - \phi(U(q^0, t, \kappa))). \end{aligned} \quad (\text{A.4})$$

Since q^1 and q^0 are fixed, so are $\phi(U(q^1, t, \kappa))$ and $\phi(U(q^0, t, \kappa))$ (equal to v and u respectively).

$$\phi(U(q, t, \kappa)) - u = \varphi^{CS}(q; t, \kappa)_U (v - u). \quad (\text{A.5})$$

$$\phi(U(q, t, \kappa)) = \varphi^{CS}(q; t, \kappa)_U a + b. \quad (\text{A.6})$$

Since $\varphi^{CS}(q; t, \kappa)_U \stackrel{c}{=} U(q, t, \kappa)$,

$$\begin{aligned} \phi(U(q, t, \kappa)) &= (dU(q, t, \kappa) + e)a + b \\ &= fU(q, t, \kappa) + g \stackrel{c}{=} U(q, t, \kappa). \end{aligned} \quad (\text{A.7})$$

Sufficiency: If $U' = aU + b$

$$\begin{aligned} \varphi^{CS}(q; t, \kappa)_{U'} &= \frac{aU(q, t, \kappa) + b - aU(q^0, t, \kappa) - b}{aU(q^1, t, \kappa) + b - aU(q^0, t, \kappa) - b} = \\ \frac{U(q, t, \kappa) - U(q^0, t, \kappa)}{U(q^1, t, \kappa) - U(q^0, t, \kappa)} &= \varphi^{CS}(q; t, \kappa)_U. \end{aligned} \quad (\text{A.8})$$

Suppose not. Then $U' \stackrel{o}{=} U$. $\varphi^{CS}(q; t, \kappa)_{U'} = \frac{\phi(U(q, t, \kappa)) - \phi(U(q^0, t, \kappa))}{\phi(U(q^1, t, \kappa)) - \phi(U(q^0, t, \kappa))} = \bar{a}\phi(U(q, t, \kappa)) + \bar{b} \stackrel{o}{=} U(q, t, \kappa)$. Since $\varphi^{CS}(q; t, \kappa)_U \stackrel{c}{=} U(q, t, \kappa)$ and $\varphi^{CS}(q; t, \kappa)_{U'} \stackrel{o}{=} U(q, t, \kappa)$, this implies $\varphi^{CS}(q; t, \kappa)_{U'} \stackrel{o}{=} \varphi^{CS}(q; t, \kappa)_U$.

b) Invariance to (q^1, q^0) .

Given (q^1, q^0) and (\bar{q}^1, \bar{q}^0) are fixed points,

$\varphi^{CS}(q; t, \kappa)_{(q^1, q^0)} = aU(q, t, \kappa) + b$ and $\varphi^{CS}(q; t, \kappa)_{(\bar{q}^1, \bar{q}^0)} = dU(q, t, \kappa) + e$. Thus, $\varphi^{CS}(q; t, \kappa)_{(q^1, q^0)} = \varphi^{CS}(q; t, \kappa)_{(\bar{q}^1, \bar{q}^0)} \leftrightarrow aU(q, t, \kappa) + b = dU(q, t, \kappa) + e \leftrightarrow b = d$ and $a = c \leftrightarrow U(q^1, t, \kappa) = U(\bar{q}^1, t, \kappa)$ and $U(q^0, t, \kappa) = U(\bar{q}^0, t, \kappa)$.

If not, $\varphi^{CS}(q; t, \kappa)_{(q^1, q^0)} = \left(\frac{a}{d}\right)\varphi^{CS}(q; t, \kappa)_{(\bar{q}^1, \bar{q}^0)} - \left(\frac{ea}{d} - b\right) \stackrel{c}{=}$
 $\varphi^{CS}(q; t, \kappa)_{(\bar{q}^1, \bar{q}^0)}$. \square

Proof of Lemma 1 (ME):

Given the assumptions in the description section,

$$\varphi^{ME}(q; t, \kappa) = \frac{U(q, t, \kappa)}{U(q^1, t, \kappa)}. \quad (\text{A.9})$$

1.a) Ordering of levels

Since q^1 is a fixed point, $U(q^1, t, \kappa) = \bar{c}$ so

$$\varphi^{ME}(q; t, \kappa) = \frac{1}{\bar{c}}U(q, t, \kappa) \stackrel{r}{=} U(q, t, \kappa), \quad (\text{A.10})$$

$$\frac{1}{\bar{c}} > 0 \leftrightarrow U(q^1, t, \kappa) > 0.$$

1.b) Ordering of Differences

Since $\varphi^{ME}(q; t, \kappa) \stackrel{r}{=} U(q, t, \kappa)$ (if $U(q^1, t, \kappa) \neq 0$), $\varphi^{ME} \in S^c(\hat{U}) \leftrightarrow U \in S^c(\hat{U})$.

2) Completeness:

Obviously, if $U(q^1, t, \kappa) \neq 0$, then any q defined for $U(q, t, \kappa)$ is defined for $\varphi^{ME}(q; t, \kappa)$, but if $U(q^1, t, \kappa) = 0$, $\varphi^{ME}(q; t, \kappa)$ is undefined for all q .

3) Uniqueness:

a) Invariance to $\phi(U)$. Let $U' = \phi(U)$.

Necessity: By the definition of uniqueness $\varphi^{ME}(q; t, \kappa)_U = \varphi^{ME}(q; t, \kappa)_{U'}$. Substituting for $\varphi^{ME}(q; t, \kappa)_{U'}$:

$$\varphi^{ME}(q; t, \kappa)_U = \frac{\phi(U(q, t, \kappa))}{\phi(U(q^1, t, \kappa))} = \frac{\phi(U(q, t, \kappa))}{\bar{c}}, \quad (\text{A.11})$$

since q^1 is fixed. Since $\varphi^{ME}(q; t, \kappa)_U \stackrel{r}{=} U(q, t, \kappa)$,

$$\phi(U(q, t, \kappa)) = \frac{a}{\bar{c}} U(q, t, \kappa) \stackrel{r}{=} U(q, t, \kappa). \quad (\text{A.12})$$

Sufficiency: If $U' = aU$,

$$\varphi^{ME}(q; t, \kappa)_{U'} = \frac{aU(q, t, \kappa)}{aU(q^1, t, \kappa)} = \frac{U(q, t, \kappa)}{U(q^1, t, \kappa)} = \varphi^{ME}(q; t, \kappa)_U. \quad (\text{A.13})$$

Suppose not. Then $U' \stackrel{o}{=} U$. $\varphi^{ME}(q; t, \kappa)_{U'} = \frac{\phi(U(q, t, \kappa))}{\phi(U(q^1, t, \kappa))} = \tilde{a}\phi(U(q, t, \kappa)) \stackrel{o}{=} U(q, t, \kappa)$. Since $\varphi^{ME}(q; t, \kappa)_U \stackrel{r}{=} U(q, t, \kappa)$, this implies $\varphi^{ME}(q; t, \kappa)_{U'} \stackrel{o}{=} \varphi^{ME}(q; t, \kappa)_U$.

b) Invariance to q^1 .

$$\begin{aligned} \varphi^{ME}(q; t, \kappa)_{q^1} &= \varphi^{ME}(q; t, \kappa)_{\bar{q}^1} \leftrightarrow \frac{U(q, t, \kappa)}{U(q^1, t, \kappa)} = \frac{U(q, t, \kappa)}{U(\bar{q}^1, t, \kappa)} \leftrightarrow U(q^1, t, \kappa) = \\ &U(\bar{q}^1, t, \kappa). \text{ Suppose not. Then } \varphi^{ME}(q; t, \kappa)_{\bar{q}^1} = \frac{U(q, t, \kappa)}{U(\bar{q}^1, t, \kappa)} = \\ &\frac{U(q^1, t, \kappa)}{U(q^1, t, \kappa)} \frac{U(q, t, \kappa)}{U(\bar{q}^1, t, \kappa)} = \frac{U(q^1, t, \kappa)}{U(\bar{q}^1, t, \kappa)} \varphi^{ME}(q; t, \kappa)_{q^1} \stackrel{c}{=} \varphi^{ME}(q; t, \kappa)_{q^1}. \text{ Thus,} \\ &\varphi^{ME}(q; t, \kappa)_{\bar{q}^1} \stackrel{r}{=} \varphi^{ME}(q; t, \kappa)_{q^1}. \square \end{aligned}$$

Proof of Lemma 1 (SG):

Given the assumptions in the description section, $\varphi^{SG}(q; t, \kappa) = p$ where p is implicitly defined

$$\tilde{U}((q, t, \kappa), 1) = \tilde{U}((q^1, t, \kappa), p, (q^0, t, \kappa), (1 - p)). \quad (\text{A.14})$$

1.a) Ordering of levels:

Solve the above equation for p . Since all arguments of the right hand side but p are fixed: $\tilde{U}((q, t, \kappa), 1) = U(q, t, \kappa) = \bar{U}(p)$. Then,

$$p = \bar{U}^{-1}(U(q, t, \kappa)). \quad (\text{A.15})$$

Since \bar{U} is an increasing monotonic function, so is its inverse. Thus, $\varphi^{SG}(q; t, \kappa) \stackrel{c}{=} U(q, t, \kappa)$.

1.b) Ordering of differences

Sufficiency: If $\tilde{U}(\cdot) \stackrel{c}{=} p\hat{U}(q, t, \kappa)$, then the problem above becomes

$$1 \cdot \hat{U}(q, t, \kappa) = \hat{U}(q, t, \kappa) = p\hat{U}(q^1, t, \kappa) + (1 - p)\hat{U}(q^0, t, \kappa), \quad (\text{A.16})$$

such that

$$p = \varphi^{SG}(q; t, \kappa) = \frac{\hat{U}(q, t, \kappa) - \hat{U}(q^0, t, \kappa)}{\hat{U}(q^1, t, \kappa) - \hat{U}(q^0, t, \kappa)}, \quad (\text{A.17})$$

which has the same structure as the CS problem. Following that proof $\varphi^{SG}(q; t, \kappa) \stackrel{c}{=} \hat{U}(q, t, \kappa)$ and $\varphi^{SG} \in S^c(\hat{U})$.

Necessity: If $\varphi^{SG}(q; t, \kappa) = p = a\hat{U}(q, t, \kappa) + b$, then $\frac{p-b}{a} = \hat{U}(q, t, \kappa)$. Given that there exists

$$\tilde{U}((q, t, \kappa), 1) = \tilde{U}((q^1, t, \kappa), p, (q^0, t, \kappa), 1 - p) \quad (\text{A.18})$$

and that there exists ϕ such that

$$\phi(\tilde{U}((q, t, \kappa), 1)) = \hat{U}(q, t, \kappa), \quad (\text{A.19})$$

which does not affect the choice of p above, then these two facts may be combined to yield

$$\frac{p-b}{a} = \phi(\tilde{U}((q^1, t, \kappa), p, (q^0, t, \kappa), 1 - p)). \quad (\text{A.20})$$

Since the left hand side is linear in p , so is the right hand side. Since the right hand side is ordinaly equivalent to \tilde{U} , then \tilde{U} must be homothetic in p (i.e. $\tilde{U}(x_{win}, p, x_{loss}, 1 - p) \stackrel{o}{=} p\hat{U}(x_{win}) + (1 - p)\hat{U}(x_{loss})$).

If $\tilde{U}(x_{win}, p, x_{loss}, 1 - p) \stackrel{o}{=} p\hat{U}(x_{win}) + (1 - p)\hat{U}(x_{loss})$, $\varphi^{SG}(q; t, \kappa) \stackrel{c}{=} \hat{U}(q, t, \kappa)$. Thus, $\varphi^{SG} \in S^c(\hat{U})$. If not, then $\varphi^{SG}(q; t, \kappa) \stackrel{o}{=} U(q, t, \kappa)$ and $\varphi^{SG} \in S^o(\hat{U})$.

2) Completeness

Domain: $\varphi^{SG}(q; t, \kappa) \in [0, 1]$ by the properties of a distribution function. By certainty equivalence, $U(q^0, t, \kappa) = \tilde{U}((q^1, t, \kappa), 0, (q^0, t, \kappa), 1)$ and $U(q^1, t, \kappa) = \tilde{U}((q^1, t, \kappa), 1, (q^0, t, \kappa), 0)$. By assumption, $\tilde{U}_p(\cdot), \tilde{U}_\xi(\cdot) \geq 0$ for all (p, ξ_k) . Then the domain is defined by $(q^1, t, \kappa) \mathcal{R}(q, t, \kappa) \mathcal{R}(q^0, t, \kappa)$ since, if $(q, t, \kappa) P(q^1, t, \kappa)$, (P indicating strict preference), it is not possible to increase p above one to achieve indifference, whereas if $(q^0, t, \kappa) P(q, t, \kappa)$, one cannot reduce p below zero.

3) Uniqueness:

a) Invariance to $\phi(\tilde{U})$. Let $\tilde{U}' = \phi(\tilde{U})$.

Sufficiency: $\varphi^{SG}(q; t, \kappa)_{\tilde{U}'} = \bar{U}^{-1}(\phi^{-1}(\phi(\tilde{U}((q, t, \kappa), 1)))) = \bar{U}^{-1}(\tilde{U}((q, t, \kappa), 1)) = \varphi^{SG}(q; t, \kappa)_{\tilde{U}}$ such that any two representations related by an increasing monotonic (i.e. with an inverse) function will yield the same result.

Necessity: If $\varphi^{SG}(q; t, \kappa)_{\tilde{U}'} = \varphi^{SG}(q; t, \kappa)_{\tilde{U}}$, then p must solve

$$\tilde{U}((q, t, \kappa), 1) = \tilde{U}((q^1, t, \kappa), p, (q^0, t, \kappa), (1 - p)).$$

and

$$\phi(\tilde{U}(q, t, \kappa), 1) = \phi(\tilde{U}((q^1, t, \kappa), p, (q^0, t, \kappa), (1 - p))). \quad (\text{A.21})$$

But the choice of p is invariant to ϕ (or ϕ^{-1}), so they are the same problem.

Invariance to $\hat{U}' \doteq \hat{U}$ if $\tilde{U}(x_{win}, p, x_{loss}, 1-p) \doteq p\hat{U}(x_{win}) + (1-p)\hat{U}(x_{loss})$.

This is a standard result of expected utility theory. If

$\tilde{U}(x_{win}, p, x_{loss}, 1-p) \doteq p\hat{U}(x_{win}) + (1-p)\hat{U}(x_{loss})$, then

$$\varphi^{SG}(q; t, \kappa) = \frac{\hat{U}(q, t, \kappa) - \hat{U}(q^0, t, \kappa)}{\hat{U}(q^1, t, \kappa) - \hat{U}(q^0, t, \kappa)}, \quad (\text{A.22})$$

which is the same as for CS, but with \hat{U} replacing U . The result follows.

b) Invariance to choice of standard gamble.

$$\varphi^{SG}(q; t, \kappa)_{(q^1, q^0)} = p :$$

$$\tilde{U}((q, t, \kappa), 1) = \tilde{U}((q^1, t, \kappa), p, (q^0, t, \kappa), (1-p)) = \bar{U}(p), \quad (\text{A.23})$$

$$\varphi^{SG}(q; t, \kappa)_{(\bar{q}^1, \bar{q}^0)} = p' :$$

$$\tilde{U}((q, t, \kappa), 1) = \tilde{U}((\bar{q}^1, t, \kappa), p, (\bar{q}^0, t, \kappa), (1-p)) = \bar{U}'(p'). \quad (\text{A.24})$$

Then

$$\begin{aligned} p = p' &\leftrightarrow \bar{U}^{-1}(U) = \bar{U}'^{-1}(U) \leftrightarrow \\ &\tilde{U}((q^1, t, \kappa), p, (q^0, t, \kappa), (1-p)) \\ &= \tilde{U}((\bar{q}^1, t, \kappa), p, (\bar{q}^0, t, \kappa), (1-p)) \quad \forall p. \end{aligned} \quad (\text{A.25})$$

Pick $p = 1$ and $p = 0$. Then the above condition reduces to two certainty equivalents: $U(q^1, t, \kappa) = U(\bar{q}^1, t, \kappa)$ and $U(q^0, t, \kappa) = U(\bar{q}^0, t, \kappa)$.

Suppose not. Then, if $\tilde{U}(x_{win}, p, x_{loss}, 1 - p) \stackrel{\circ}{=} p\hat{U}(x_{win}) + (1 - p)\hat{U}(x_{loss})$,

$$\varphi_{(q^1, q^0)}^{SG}(q; t, \kappa) = \frac{\hat{U}(q, t, \kappa) - \hat{U}(q^0, t, \kappa)}{\hat{U}(q^1, t, \kappa) - \hat{U}(q^0, t, \kappa)} = a\hat{U}(q, t, \kappa) + b, \quad (\text{A.26})$$

$$\varphi_{(\bar{q}^1, \bar{q}^0)}^{SG}(q; t, \kappa) = \frac{\hat{U}(q, t, \kappa) - \hat{U}(\bar{q}^0, t, \kappa)}{\hat{U}(\bar{q}^1, t, \kappa) - \hat{U}(\bar{q}^0, t, \kappa)} = a'\hat{U}(q, t, \kappa) + b', \quad (\text{A.27})$$

which is the same as the CS case, with \hat{U} replacing U . Thus,

$$\varphi_{(q^1, q^0)}^{SG}(q; t, \kappa) \stackrel{\circ}{=} \varphi_{(\bar{q}^1, \bar{q}^0)}^{SG}(q; t, \kappa). \quad (\text{A.28})$$

If $\tilde{U}(x_{win}, p, x_{loss}, 1 - p) \not\stackrel{\circ}{=} p\hat{U}(x_{win}) + (1 - p)\hat{U}(x_{loss})$,

$$\varphi^{SG}(q; t, \kappa)_{(q^1, q^0)} = \bar{U}^{-1}(\tilde{U}((q, t, \kappa), 1)), \quad (\text{A.29})$$

$$\varphi^{SG}(q; t, \kappa)_{(\bar{q}^1, \bar{q}^0)} = \bar{U}'^{-1}(\tilde{U}'((q, t, \kappa), 1)), \quad (\text{A.30})$$

$$\begin{aligned} \varphi^{SG}(q; t, \kappa)_{(q^1, q^0)} &= \\ \bar{U}^{-1}(\bar{U}'(\varphi^{SG}(q; t, \kappa)_{(\bar{q}^1, \bar{q}^0)})) &\stackrel{\circ}{=} \varphi^{SG}(q; t, \kappa)_{(\bar{q}^1, \bar{q}^0)}. \end{aligned} \quad (\text{A.31})$$

□

Proof of Lemma 1 (TTO):

Given the assumptions in the description section, $\varphi^{TTO}(q; t, \kappa) = (t - m)/t$ where m is implicitly defined:

$$U(q, t, \kappa) = U(q^1, t - m, \kappa). \quad (\text{A.32})$$

1.a) Ordering levels:

Solve the above equation for m . Since all arguments of the right hand side but m are fixed: $U(q, t, \kappa) = \bar{U}(t - m)$. Then

$$t - m = \bar{U}^{-1}(U(q, t, \kappa)). \quad (\text{A.33})$$

Since \bar{U} is an increasing monotonic function, so is its inverse. Thus, $t - m \stackrel{\circ}{=} U(q, t, \kappa)$. But $\varphi^{TTO}(q; t, \kappa) = (t - m)/t = (1/t)\bar{U}^{-1}(U(q, t, \kappa))$ which is not an allowable transformation if t varies since it would depend on an argument of the function itself. If t is fixed, this is simply a ratio scale transform of an allowable transform, which is itself allowable.

Ordering levels of morbidity

Sufficiency: If $U(\cdot) \stackrel{\circ}{=} t\mu(q, \kappa)$, then the problem above becomes

$$\phi(t\mu(q, \kappa)) = \phi((t - m)\mu(q^1, \kappa)), \quad (\text{A.34})$$

such that $\varphi^{TTO}(q; t, \kappa) = \frac{t-m}{m} = \frac{\mu(q, \kappa)}{\mu(q^1, \kappa)} = a\mu(q, \kappa) \stackrel{r}{=} \mu(q, \kappa)$ since q^1 is fixed such that $\mu(q^1, \kappa) = \frac{1}{a} > 0$.

Necessity: If $\varphi^{TTO}(q; t, \kappa) \stackrel{r}{=} \mu(q, \kappa)$, $\varphi^{TTO}(q; t, \kappa)$ must be independent of t since the right hand side of the equation is independent of t . By the definition of independence, this means that $\varphi^{TTO}(q; t, \kappa)$ must be the same for all values of t , i.e.

$$U(q, \lambda t, \kappa) = U(q^1, \lambda(t - m), \kappa) \forall \lambda > 0. \quad (\text{A.35})$$

Set $\lambda = 1/t$. Then $U(q, 1, \kappa) = U(q^1, (t - m)/t, \kappa) = \bar{U}((t - m)/t)$. But then all arguments but q are fixed on the left hand side so $\bar{U}(q, \kappa) = \bar{U}((t - m)/t)$ such that $(t - m)/t = \bar{U}^{-1}(\bar{U}(q, \kappa))$. Let $\bar{U}^{-1}(\bar{U}) = \mu$. Then rearrangement of the above expression yields $t - m = t\mu(q, \kappa)$. From above, $t - m \stackrel{\circ}{=} U$ so $U(q, t, \kappa) \stackrel{\circ}{=} t\mu(q, \kappa)$.

Suppose not. Then $\frac{t-m}{t} = f(t) \neq \phi(\mu(q, \kappa))$.

1.b) Ordering differences

If $U(x) \stackrel{\circ}{=} t\mu(q, t, \kappa)$, $\varphi^{TTO}(q; t, \kappa) \stackrel{\circ}{=} \mu(q, \kappa)$. Thus, $\varphi^{TTO} \in S^c(\mu)$. If $\hat{U}(x) \stackrel{\circ}{=} \mu(q, \kappa)t$, then $\varphi^{TTO} \in S^c(\hat{U})$ for any choice of t . If $\hat{U}(x) \not\stackrel{\circ}{=} \mu(q, \kappa)t$,

then $\varphi^{TTO}(q; t, \kappa) \stackrel{o}{=} U(q, t, \kappa)$ if and only if t is fixed and then $\varphi^{TTO} \in S^o(\hat{U})$.

2) Completeness

Domain: Since $U_q(x), U_t(x) \geq 0$ and $U(q^0, t, \kappa) = U(q, 0, \kappa)$ by assumption, $(q, t, \kappa) \mathfrak{R}(q^0, t, \kappa)$ for all (t, κ) since $t - m \geq 0$ (i.e. cannot increase m to achieve indifference) and $(q^1, t, \kappa) \mathfrak{R}(q, t, \kappa)$ for all (t, κ) since $m \geq 0$ (cannot reduce m to achieve indifference). For $q \notin [q^0, q^1]$, it is not possible to adjust $t - m$ to achieve indifference.

3) Uniqueness:

a) Invariance to $\phi(U)$. Let $U' = \phi(U)$.

Sufficiency: $\varphi^{TTO}(q; t, \kappa)_{U'}$ is defined by $\bar{U}^{-1}(\phi^{-1}(\phi(U(q, t, \kappa)))) = \bar{U}^{-1}(U(q, t, \kappa)) = \varphi^{TTO}(q; t, \kappa)_U$, such that any two representations related by an increasing monotonic (i.e. with an inverse) function will yield the same result.

Necessity: If $\varphi^{TTO}(q; t, \kappa)_{U'} = \varphi^{TTO}(q; t, \kappa)_U$, then $t - m$ must solve

$$U(q, t, \kappa) = U(q^1, t - m, \kappa), \quad (\text{A.36})$$

and

$$\phi(U(q, t, \kappa)) = \phi(U(q^1, t - m, \kappa)). \quad (\text{A.37})$$

But the choice of $t - m$ is invariant to ϕ (or ϕ^{-1}), so they are the same problem.

Invariance to $\mu' \stackrel{o}{=} \mu$ if $U(x) \stackrel{o}{=} t\mu(q, \kappa)$.

If $U(x) \stackrel{o}{=} t\mu(q, \kappa)$, then

$$\varphi^{TTO}(q; t, \kappa) = \frac{\mu(q, \kappa)}{\mu(q^1, \kappa)}, \quad (\text{A.38})$$

which is the same as for ME, but with μ replacing U . The result follows.

b) Invariance to choice of reference state.

Since,

$$\varphi^{TTO}(q; t, \kappa)_{q^1} = \bar{U}^{-1}(U(q, t, \kappa))/t, \quad (\text{A.39})$$

$$\varphi^{TTO}(q; t, \kappa)_{\bar{q}'} = \tilde{U}^{-1}(U(q, t, \kappa))/t, \quad (\text{A.40})$$

then

$$\bar{U}^{-1}(U) = \tilde{U}^{-1}(U) \leftrightarrow \bar{U} = \tilde{U} \leftrightarrow U(q^1, t, \kappa) = U(\bar{q}', t, \kappa). \quad (\text{A.41})$$

for all (t, κ) .

Suppose not. Then, if $U(x) \stackrel{\circ}{=} t\mu(q, \kappa)$, this is the same problem as ME and the result follows. If $U(x) \stackrel{\circ}{\neq} t\mu(q, \kappa)$, then

$$\varphi^{TTO}(q; t, \kappa)_{q^1} = \bar{U}^{-1}(\tilde{U}(\varphi^{TTO}(q; t, \kappa)_{\bar{q}'})) \stackrel{\circ}{=} \varphi^{TTO}(q; t, \kappa)_{\bar{q}'}. \quad (\text{A.42})$$

□

Proof of Lemma 1 (ES):

Given the assumptions in the description section, $\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) = (t_j - m)/t_i$ where m is implicitly defined

$$U(q_i, t_i, \kappa_i) = U(q^j, t_j - m, \kappa_j). \quad (\text{A.43})$$

1.a) Ordering levels:

Recognizing that all arguments but m are fixed on the right hand side, solve the above equation for m : $U(q_i, t_i, \kappa_i) = \bar{U}^j(t_j - m)$. Then

$$t_j - m = \bar{U}^{j-1}(U(q_i, t_i, \kappa_i)). \quad (\text{A.44})$$

Since \bar{U}^j is an increasing monotonic function, so is its inverse. Thus, $t_j - m \stackrel{o}{=} U(q_i, t_i, \kappa_i)$. But $\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) = (t_j - m)/t_i = (1/t_i)\bar{U}^{j-1}(U(q_i, t_i, \kappa_i))$ which is not an allowable transformation if t_i varies since it would depend on an argument of the function itself. If t_i is fixed, this is simply a ratio scale transform of an allowable transform, which is itself allowable.

Ordering levels of morbidity:

Sufficiency: If $U(x_i) \stackrel{o}{=} t_i \mu(q_i, \kappa_i)$, then, since q_j is fixed and $\mu(q_j, \kappa_j) > 0$, the proof from TTO may be applied.

Necessity: If $\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) \stackrel{r}{=} \mu(q_i, \kappa_i)$, $\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j)$ must be independent of t_i since the right hand side of the equation is independent of t_i . By the definition of independence, this means that $\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j)$ must be the same for all values of t_i , i.e.

$$U(q_i, \lambda t_i, \kappa_i) = U(q_j, \lambda(t_j - m), \kappa_j) \quad \forall \lambda > 0. \quad (\text{A.45})$$

Set $\lambda = 1/t_i$ and the proof for TTO may be applied.

1.b) Ordering differences

If $U(x_i) \stackrel{o}{=} t_i \mu(q_i, \kappa_i)$ for all i , $\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) \stackrel{c}{=} \mu(q_i, \kappa_i)$. Thus, $\varphi^{ES} \in S^c(\mu^i)$. If not, then $\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) \stackrel{o}{=} U(q_i, t_i, \kappa_i)$ and $\varphi^{ES} \in S^o(U^i)$ if and only if t_i is fixed.

2) Completeness

Domain: Since $U_\xi^i(x_i), U_t^i(x_i) \geq 0$ for all ξ_k, t by assumption, and $t_j - m \in [0, t_j]$, then $(q_j, t_j, \kappa_j) \mathcal{R}(q_i, t_i, \kappa_i) \mathcal{R}(q_j, 0, \kappa_j)$ since if state i is preferred to j at t_j , one cannot increase t_j to achieve indifference, whereas if state j at $t_j = 0$ is preferred to state i , one cannot decrease t_j to achieve indifference.

3) Uniqueness:

a) Invariance to $\phi(U)$. Let $U' = \phi(U)$.

$\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j)_U$ is defined by $U(q_i, t_i, \kappa_i) = U(q_j, t_j - m, \kappa_j)$ and

$\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j)_{U'}$ is defined by $U'(q_i, t_i, \kappa_i) = U'(q_j, t_j - m, \kappa_j)$.

But this is the same problem as for TTO, except the arguments of the right hand side function have changed. Since that proof is unaffected by changes in the argument values, the result carries through.

Invariance to $\mu' \stackrel{c}{=} \mu$ if $U(x_i) \stackrel{o}{=} t_i \mu(q_i, \kappa_i)$.

Again, this problem is the same as for TTO with changes in the values of the arguments which do not affect the proofs. The result follows.

b) Invariance to choice of reference state

Again, this is the same problem as for TTO, although the dimensionality of the reference state has increased by the inclusion of personal characteristics. \square

Proof of Lemma 1 (PE):

Given the assumptions in the description section,

$\varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) = (N - m)/N$ where m is implicitly defined

$$W(q_1, \dots, q_N, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) =$$

$$W(q_1^1, \dots, q_{N-m}^1, q_{N-m+1}^0, \dots, q_N^0, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N), \quad (\text{A.46})$$

which may be rewritten as

$$\tilde{W}(\{q\}, N) = \tilde{W}(\{q^1\}, N - m, \{q^0\}, m). \quad (\text{A.47})$$

1.a) Ordering levels of societal health

Since $q^1, q^0, N, t_i, \kappa_i$ are fixed in the expression above, the right hand side may be expressed

$$= \bar{W}(m). \quad (\text{A.48})$$

Taking the inverse of \bar{W} on both sides

$$m = \bar{W}^{-1}(\bar{W}(\{q\}, N)) \stackrel{\circ}{=} \tilde{W}(\{q\}, N), \quad (\text{A.49})$$

since \bar{W} is monotonic (in order to have an inverse).

But $\varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) = (N - m)/N = 1 - (m/N) = 1 - (1/N)(\bar{W}^{-1}(\bar{W}(\cdot)))$ which is not an allowable transformation if N varies since it then depends on an argument of the function itself. If N is fixed, this is simply a ratio scale transform of an allowable transform, which is itself allowable.

1.b) Ordering outcomes

Sufficiency: If $W(\cdot) \stackrel{\circ}{=} \sum_{i=1}^N U(q_i, t_i, \kappa_i)$ and $t_i = t, \kappa_i = \kappa$ for all i , then the problem above becomes

$$U(q_i, t_i, \kappa_i)N = U(q^1, t_i, \kappa_i)(N - m) + U(q^0, t_i, \kappa_i)m, \quad (\text{A.50})$$

$$\begin{aligned} \varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) &= (N - m)/N \\ &= \frac{U(q_i, t_i, \kappa_i) - U(q^0, t_i, \kappa_i)}{U(q^1, t_i, \kappa_i) - U(q^0, t_i, \kappa_i)}. \end{aligned} \quad (\text{A.51})$$

Since q^1, q^0, t_i, κ_i are fixed,

$\varphi^{PE}(q_1, \dots, q_N, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) = a(U(q_i, t_i, \kappa_i)) + b \stackrel{\circ}{=} U(q_i, t_i, \kappa_i)$, where $a > 0$ since $(q^1, t_i, \kappa_i) \mathfrak{R}(q^0, t_i, \kappa_i)$. If $W(\cdot) \stackrel{\circ}{=} \sum U(q_i, t_i, \kappa_i)$ and $t_i \neq t_j$ or $\kappa_i \neq \kappa_j$ and $U(q_i, t_i, \kappa_i) \stackrel{\circ}{=} v(q) + b(t_i, \kappa_i) + c$ (i.e. $U_{q,\S} = 0$), then the problem above becomes

$$Nv(q) + \sum_{i=1}^N b(t_i, \kappa_i) = (N - m)v(q^1) + mv(q^0) + \sum_{i=1}^N b(t_i, \kappa_i), \quad (\text{A.52})$$

such that $\varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) = \frac{v(q) - v(q^0)}{v(q^1) - v(q^0)}$ which, since $q^0, q^1, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N$ are fixed, is cardinally related to U .

Necessity: Consider the restriction on the SWF. If $W \neq \sum_{i=1}^N U(q_i, t_i, \kappa_i)$, then the independence from t results of the TTO section may be applied here (with respect to N): then $\varphi^{PE} \neq U$.

Consider the restriction on (t_i, κ_i) . If these are allowed to vary across i , then the problem becomes

$$\sum_{i=1}^N (U(q, t_i, \kappa_i)) = \sum_{i=1}^{N-m} (U(q^1, t_i, \kappa_i)) + \sum_{i=N-m+1}^N U(q^0, t_i, \kappa_i). \quad (\text{A.53})$$

Standardizing to a given (t_i, κ_i) :

$$\begin{aligned} & NU(q, \bar{t}, \bar{\kappa}) + \sum_{i=1}^N \sum_{\S} U_{\S}(q, \bar{t}, \bar{\kappa}) d\S_i = \\ & (N-m)(U(q^1, \bar{t}, \bar{\kappa})) + m(U(q^0, \bar{t}, \bar{\kappa})) + \\ & \sum_{i=1}^{N-m} \sum_{\S} U_{\S}(q^1, \bar{t}, \bar{\kappa}) d\S_i + \sum_{i=N-m+1}^N \sum_{\S} U_{\S}(q^0, \bar{t}, \bar{\kappa}) d\S_i. \end{aligned} \quad (\text{A.54})$$

Solving

$$\begin{aligned} \varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) &= \frac{U(q, \bar{t}, \bar{\kappa}) - U(q^0, \bar{t}, \bar{\kappa})}{U(q^1, \bar{t}, \bar{\kappa}) - U(q^0, \bar{t}, \bar{\kappa})} \\ &+ \left(\frac{1}{N} \right) \left[\frac{\sum_{i=1}^{N-m} \sum_{\xi} \sum_{\S} U_{\xi, \S}(q, \bar{t}, \bar{\kappa}) d\S_i (q^1 - q)}{U(q^1, \bar{t}, \bar{\kappa}) - U(q^0, \bar{t}, \bar{\kappa})} \right. \\ &\left. + \frac{\sum_{i=N-m+1}^N \sum_{\xi} \sum_{\S} U_{\xi, \S}(q, \bar{t}, \bar{\kappa}) d\S_i (q^0 - q)}{U(q^1, \bar{t}, \bar{\kappa}) - U(q^0, \bar{t}, \bar{\kappa})} \right], \end{aligned} \quad (\text{A.55})$$

which $= f(U(q), q)$ unless $U_{\xi, \S} = 0$ for all ξ, \S , or $dx = 0$. The latter condition is the identical agents condition. The former condition requires $U(q, t, \kappa) \stackrel{c}{=} v(q) + b(t, \kappa)$.

Ordering expected outcomes

Sufficiency: Let N be large. Define P_I as the proportion of agents in the sample with characteristics (t_i, κ_i) . Since the selection rule is random and N is large, $P_I^N = P_I^{N-m} = P_I^m$. For an additive SWF, this implies $W(\cdot) \stackrel{\circ}{=} \sum_I P_I U(q, t_i, \kappa_i) N$, such that

$$\begin{aligned} \sum_I P_I U(q, t_i, \kappa_i) N &= \sum_I P_I U(q^1, t_i, \kappa_i) (N - m) \\ &\quad + \sum_I P_I U(q^0, t_i, \kappa_i) m \end{aligned} \quad (\text{A.56})$$

$$\begin{aligned} \varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) &= (N - m)/N = \\ &= \frac{\sum_I P_I U(q, t_i, \kappa_i) - \sum_I P_I U(q^0, t_i, \kappa_i)}{\sum_I P_I U(q^1, t_i, \kappa_i) - \sum_I P_I U(q^0, t_i, \kappa_i)}, \end{aligned} \quad (\text{A.57})$$

$$\begin{aligned} \varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) &= \\ &= a \sum_I P_I U(q, t_i, \kappa_i) + b \stackrel{c}{=} E(U(q, t_i, \kappa_i)). \end{aligned} \quad (\text{A.58})$$

Necessity: Suppose (i) N is small or (ii) the selection rule (represented by S) is non-random. Then $P_I^N = f(t_1, \dots, t_N, \kappa_1, \dots, \kappa_N)$,

$$P_I^{N-m} = \bar{f}(t_1, \dots, t_{N-m}, \kappa_1, \dots, \kappa_{N-m}),$$

$P_I^m = \hat{f}(t_{N-m+1}, \dots, t_N, \kappa_{N-m+1}, \dots, \kappa_N)$, and $P_I^N \neq P_I^{N-m} \neq P_I^m$. Then

$$\begin{aligned} \varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) &= \\ &= a(t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) \sum_I P_I U(q, t_i, \kappa_i) + b(t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) \\ &\quad \stackrel{c}{\neq} E(U(q, t_i, \kappa_i)), \end{aligned} \quad (\text{A.59})$$

since the transform depends on arguments of the function itself.

1.c) Ordering differences

If $W(\cdot) \stackrel{\circ}{=} \sum_{i=1}^N U(x_i) = \sum_{i=1}^N v^i(q)$ and $v^i(q) = v(q) \forall i$ or $U_{\xi, \S}^i = 0 \forall (\xi, \S)$,

then from exactness $\varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) \stackrel{c}{=} U(q_i, t_i, \kappa_i)$, such

that $\varphi^{PE} \in S^c(U)$. If $U \in S^c(\hat{U})$, then $\varphi^{PE} \in S^c(\hat{U})$. If not, but N and $(t_1, \dots, t_N, \kappa_1, \dots, \kappa_N)$ are fixed $\varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) \stackrel{o}{=} W(U(q_1, t_1, \kappa_1), \dots, U(q_N, t_N, \kappa_N))$, such that $\varphi^{PE} \in S^o(U)$.

2) Completeness

Domain: Follow the same argument used in TTO, but replace R with R_{social} , q_t with q_i and t with N . Since $\bar{W}_q, \bar{W}_N > 0$ and $N - m \in [0, N]$, the result follows.

3) Uniqueness:

a) Invariance to $\phi(W)$. Let $W' = \phi(W)$.

Then this is the same problem as for TTO but W replaces U .

Invariance to $\phi^i(U^i)$. Let $U^{i'} = \phi^i(U^i)$.

Because the SWF involves interpersonal comparisons (except in the case of dictatorship), and individual specific transforms affect these comparisons, φ^{PE} will not be unique to individual specific transforms (this is a basic result of Arrow's theorem) unless the ordering is somehow flawed. When the restrictions on the ordering are relaxed, the set of allowable transforms will depend on the nature of W . If W is homothetic, \tilde{U} must be cardinally unit comparable, i.e. $\tilde{U} = aU + b^i$. If W is based on levels rather than differences, then \tilde{U} must be cardinally fully comparable, i.e. $\tilde{U} = U$. See Sen (1986) for proofs of these and similar propositions.

b) Invariance to $\{(t_i, \kappa_i)\}$ and choice of selection rule.

Since there are a finite number of agents, there exists $U(q, t_i, \kappa_i) = U^i(q)$, such that (t_i, κ_i) may be treated as an agent specific transform.

The results of (a) above then apply, (t_i, κ_i) allowed to affect utility

over q only as individual specific transforms are allowed above, and common characteristics allowed to affect utility over q only as common transforms are allowed above.

A change in selection rule may be viewed as a change in the individual specific transforms of the cured and the deceased such that (a) may be applied once again.

c) Invariance to choice of reference points (q^1, q^0).

Again, this is the same problem as for TTO, with W and N replacing U and t . \square

Proof of Lemma 2 (CS):

Sufficiency: If $U(x) = av(q)w(t, \kappa) + az(t, \kappa) + b$, then $\varphi^{CS}(q; t, \kappa) = \frac{v(q) - v(q^0)}{v(q^1) - v(q^0)} = \bar{\varphi}^{CS}(q) \forall (t, \kappa)$.

Necessity: If $\varphi^{CS}(q; t, \kappa)$ is independent of (t, κ) , then, by the definition of independence, $\varphi^{CS}(q; t, \kappa) = \bar{\varphi}^{CS}(q)$. $\varphi^{CS}(q; t, \kappa) = \frac{U(q, t, \kappa) - U(q^0, t, \kappa)}{U(q^1, t, \kappa) - U(q^0, t, \kappa)}$, but since q^1 and q^0 are fixed, this may be expressed $\varphi^{CS}(q; t, \kappa) = \frac{U(q, t, \kappa) - z(t, \kappa)}{y(t, \kappa) - z(t, \kappa)} = \frac{U(q, t, \kappa) - z(t, \kappa)}{w(t, \kappa)}$. Substituting this into the independence condition: $\bar{\varphi}^{CS}(q) = \frac{U(q, t, \kappa) - z(t, \kappa)}{w(t, \kappa)}$. Let $\bar{\varphi}^{CS}(q) = v(q)$ and rearrange: $U(q, t, \kappa) = v(q)w(t, \kappa) + z(t, \kappa)$. Since $\varphi^{CS}(q; t, \kappa)$ is invariant to affine transforms of U (see Lemma 1), this completes the proof. \square

Proof of Lemma 2 (ME):

Sufficiency: If $U(x) = av(q)w(t, \kappa)$, then $\varphi^{ME}(q; t, \kappa) = \frac{av(q)w(t, \kappa)}{av(q^1)w(t, \kappa)} = \frac{v(q)}{v(q^1)} = \bar{\varphi}^{ME}(q) \forall (t, \kappa)$.

Necessity: Independence requires $\varphi^{ME}(q; t, \kappa) = \bar{\varphi}^{ME}(q)$. Substituting the expression for $\varphi^{ME}(q; t, \kappa)$ and setting $\bar{\varphi}^{ME}(q) = v(q)$ and $U(q^1, t, \kappa) =$

$w(t, \kappa)$ (since q^1 is fixed) yields $U(q, t, \kappa) = v(q)w(t, \kappa)$. Invariance to ratio scale transforms (Lemma 1 (ME)) completes the proof. \square

Proof of Lemma 2 (SG): If the von Neumann-Morgenstern axioms hold and (t, κ) are fixed for all states of the world, then $\varphi^{SG}(q; t, \kappa) = \frac{\bar{U}(q, t, \kappa) - \bar{U}(q_{loss}, t, \kappa)}{\bar{U}(q_{win}, t, \kappa) - \bar{U}(q_{loss}, t, \kappa)}$, which is the same expression as for the CS instrument, except the restrictions derived must apply to \hat{U} and not an arbitrary U . \square

Proof of Lemma 2 (TTO):

a) Independence from κ

Sufficiency: If $U(x) = U(v(q, t), \kappa)$, then, for any κ there exists some \bar{U} such that $U(x) = \bar{U}(v(q, t))$. By Lemma 1 (TTO), TTO is invariant to transforms of U .

Necessity: Since $\varphi^{TTO}(q; t, \kappa) = \frac{t-m}{t}$, $\frac{\partial \varphi^{TTO}(q; t, \kappa)}{\partial \xi} = (-\frac{\partial m}{\partial \xi}) \frac{1}{t}$. From the expression for TTO, $U(q, t, \kappa) = U(q^1, t - m, \kappa)$. Totally differentiate this expression to get $\sum_{\xi} U_{\xi}(x) d\xi_k + U_t(x) dt = 0$ which yields $dt = -m = -\frac{\sum_{\xi} U_{\xi}(x)}{U_t(x)} d\xi_k$. Then $\frac{\partial \varphi^{TTO}(q; t, \kappa)}{\partial \xi} = -\frac{\partial \sum_{\xi} U_{\xi}(x)/U_t(x)}{\partial \xi} (\frac{\sum_{\xi} d\xi_k}{t})$. Thus, $sign(\frac{\partial \varphi^{TTO}(q; t, \kappa)}{\partial \xi}) = -sign(\sum_{\xi} \frac{\partial U_{\xi}(x)/U_t(x)}{\partial \xi}) = -sign(\sum_{\xi} \frac{\partial MRS_{\xi, t}}{\partial \xi})$.

b) Independence from t

Sufficiency: If $U \stackrel{o}{=} t\mu(q, \kappa)$, then TTO is defined by $\phi(\mu(q, \kappa)t) = \phi(\mu(q^1, \kappa)(t - m))$ such that $\varphi^{TTO}(q; t, \kappa) = \frac{t-m}{t} = \frac{\mu(q, \kappa)}{\mu(q^1, \kappa)}$.

Necessity: see Lemma 1 (TTO).

If not, then

$$\frac{\partial \varphi^{TTO}(q; t, \kappa)}{\partial t} = \left(\frac{-\partial m / \partial t}{t} - \frac{m}{t^2} \right). \quad (\text{A.60})$$

Totally differentiate the basic TTO expression to obtain

$$U_t(q, t, \kappa) dt = U_t(q^1, t - m, \kappa) dt - U_t(q^1, t - m, \kappa) dm \quad (\text{A.61})$$

$$\frac{dm}{dt} = 1 - \frac{U_t(q, t, \kappa)}{U_t(q^1, t - m, \kappa)}. \quad (\text{A.62})$$

Substituting this into the above expression yields

$$\frac{\partial \varphi^{TTO}(q; t, \kappa)}{\partial t} = \frac{1}{t} \left(\frac{U_t(q, t, \kappa)}{U_t(q^1, t - m, \kappa)} - \varphi^{TTO}(q; t, \kappa) \right) > / < 0. \quad (\text{A.63})$$

Note that if $U_t(q, t, \kappa) > U_t(q^1, t - m, \kappa)$, $\frac{\partial \varphi^{TTO}(q; t, \kappa)}{\partial t} > 0$ since $\varphi^{TTO}(q; t, \kappa) \leq 1$ for all q . \square

Proof of Lemma 2 (ES):

a) Independence from κ

If $\kappa_i \neq \kappa_j$, let $\kappa_j = \kappa_i + \delta_j$. Then $U(q_i, t_i, \kappa_i) = U(q_j, t_j - m, \kappa_i + \delta_j)$. Differentiate this expression to obtain $\sum_{\xi} U_{\xi}(x_i) d\xi_k + U_t(x_i) dt + \sum_{\S} U_{\S}(x_i) \delta_j = 0$. Solving, obtain $\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) = (1 - \frac{\sum_{\xi} U_{\xi}(x_i) d\xi_k + U_t(x_i) dt}{U_t(x_i) t_i})$. Since $\delta_j > 0$ signifies $\kappa_j > \kappa_i$, whereas $\delta_j < 0$ signifies $\kappa_j < \kappa_i$, $\frac{\partial \varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j)}{\partial \S_j} = (-\frac{\partial m}{\partial \S_j}) \frac{1}{t_i} = \frac{-U_{\S}(\cdot)}{t U_t(\cdot)} > / < 0 \leftrightarrow U_{\S} < / > 0$ and $\frac{\partial \varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j)}{\partial \S_i} > / < 0 \leftrightarrow U_{\S} > / < 0$.

b) Independence from t

If $t_i \neq t_j$, then the result depends on which time frame differs. Let t_j change. Then the effect on the ES may be found by differentiation of the ES problem: $0 = U_t(x) dt_j + U_t(x)(-dm)$, which implies $dt_j = dm$ since $U_t(x) = U_t(x)$. Then, $\frac{\partial \varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j)}{\partial t_j} = \frac{1}{t_i} (dt_j - dm) = 0$.

Let t_i change. The differentiation exercise becomes $U_t(x_i) dt_i = -U_t(x_j) dm$. Then $\frac{\partial \varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j)}{\partial t_i} = \frac{-1}{t_i} (\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) - \frac{dm}{dt_i}) = \frac{1}{t_i} (-\varphi^{ES}(q; t_i, \kappa_i, t_j, \kappa_j) + \frac{U_t(x_i)}{U_t(x_j)}) = U_t^j (U_t(x_i)(t_i) - U_t(x_j)(t_j - m)) > / < 0$ as the utility of time for i is greater/less than the utility of time for j . \square

Proof of Lemma 2 (PE):

a) If $t_i = t, \kappa_i = \kappa \forall i \in [1, N]$, but $t_i \neq t'_i$ or $\kappa_i \neq \kappa'_i$, the PE problem becomes

$\tilde{W}(q_1, \dots, q_N, t, \kappa) = \tilde{W}(q_1^1, \dots, q_{N-m}^1, q_{N-m+1}^0, \dots, q_N^0, t, \kappa)$. For any fixed N , this has the same structure as Lemma 2 (SG), with \tilde{W} and N replacing U and q respectively, and the proof follows the same structure.

Suppose the SWF is welfarist (such that N is separable from (q, t, κ)) and that utility over q depends on (t, κ) .

Then $W(\cdot) \stackrel{o}{=} \sum_{i=1}^N U(q, t_i, \kappa_i)$ and

$\varphi^{PE}(q_1, \dots, q_N; t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) = \frac{U(q, t, \kappa) - U(q^0, t, \kappa)}{U(q^1, t, \kappa) - U(q^0, t, \kappa)}$. But this is the same as Lemma 2 (CS) and the results follow from that proof.

b. Suppose $t_i \neq t_j$ or $\kappa_i \neq \kappa_j$. Then the PE problem is to find m such that $W(q_1, \dots, q_N, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) = W(q_1^1, \dots, q_{N-m}^1, q_{N-m+1}^1, \dots, q_N^1, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N)$. But this is essentially the same problem as the identical agents case except $(t_1, \dots, t_N, \kappa_1, \dots, \kappa_N)$ is an $2N$ -dimensional vector instead of a two-dimensional vector. The same argument applies over the vector.

Consider when the SWF is welfarist and utility over q depends on (t_i, κ_i) . Then $\varphi^{PE}(q_1, \dots, q_N, t_1, \dots, t_N, \kappa_1, \dots, \kappa_N) = \frac{\sum_{i=1}^N U(q, t_i, \kappa_i) - \sum_{i=1}^N U(q^0, t_i, \kappa_i)}{\sum_{i=1}^N U(q^1, t_i, \kappa_i) - \sum_{i=1}^N U(q^0, t_i, \kappa_i)} = \frac{a}{b}$. Then $\varphi_{\S_i}^{PE}(\cdot) = (U_{\S}(q, t_i, \kappa_i) - U_{\S}(q^0, t_i, \kappa_i)) - \varphi^{PE}(\cdot)(U_{\S}(q^1, t_i, \kappa_i) - U_{\S}(q^0, t_i, \kappa_i))/b$, $b > 0$ since $(q^1, t_i, \kappa_i) \mathcal{R}(q^0, t_i, \kappa_i)$. Then $\varphi_{\S_i}^{PE}(\cdot) = 0 \forall q \leftrightarrow U_{\S, \xi}(x_i) = 0 \leftrightarrow U(x_i) \stackrel{c}{=} v(q) + b(t_i, \kappa_i)$. Otherwise, reference must be made to the above expression. \square

Proof of Proposition 1:

Instruments generate identical values if and only if they are identical transformations of identical utility functions. Given reference points that fit all instruments to the same interval, this requires all transformations have the same curvature at all

points (i.e. linearity which requires the respective metrics to have constant marginal utility) and that the orderings not be influenced by factors peculiar to each instrument (i.e. independence). Thus, the proposition may be proved by combining the independence results across instruments:

- (a) intersection of independence conditions across all six instruments yields the result.
- (b) intersection of independence conditions across SG, TTO, ES, and PE yields the result.
- (c) intersection of independence conditions across SG, TTO, and ES yields the result.
- (d) intersection of independence conditions across ES and PE yields the result. \square

Proof of Lemma 3

Necessity: If the QALY satisfies WDI, then utility must satisfy WDI.

If the QALY satisfies WDI,

$$\frac{\partial(\varphi_{\xi_i}(q; t, \kappa)/\varphi_{\xi_j}(q; t, \kappa))}{\partial \xi_k} \geq 1 - (\varphi_{\xi_i}(q; t, \kappa)/\varphi_{\xi_j}(q; t, \kappa)) \quad (\text{A.64})$$

(subscripts denote partial derivatives) when $\varphi_{\xi_i} \geq \varphi_{\xi_j}$ (see derivation under axiomatic methods in the text). If $\varphi(q; t, \kappa) = \phi(U(q, t, \kappa))$ (as is always the case for CS, SG, and PE, and is the case when t is fixed for TTO and ES (see equations 2.3 to 2.15)), then (A.64) translates to

$$\frac{\partial(\phi_U U_{\xi_i}(q, t, \kappa)/\phi_U U_{\xi_j}(q, t, \kappa))}{\partial \xi_k} \geq 1 - (\phi_U U_{\xi_i}(q, t, \kappa)/\phi_U U_{\xi_j}(q, t, \kappa)) \quad (\text{A.65})$$

which reduces to

$$\frac{\partial(U_{\xi_i}(q, t, \kappa)/U_{\xi_j}(q, t, \kappa))}{\partial \xi_k} \geq 1 - (U_{\xi_i}(q, t, \kappa)/U_{\xi_j}(q, t, \kappa)) \quad (\text{A.66})$$

which is the condition for all ξ_k to be WDI in U . If U is homothetic in t , then $\varphi(q; t, \kappa) = \phi(\mu(q, \kappa))$ (for TTO and ES), then the same argument applies to μ as to U and the result is that all ξ_k are WDI in μ , which is equivalent to saying all ξ_k are WDI in U , since WDI is transformation independent.

Sufficiency: If utility satisfies WDI, then the QALY satisfies WDI.

The above argument holds in reverse since all steps are if and only if. \square

Proof of Lemma 4

Necessity: If the QALY satisfies MPI, then utility must satisfy MPI.

If the QALY satisfies MPI,

$$\frac{\partial(\varphi_{\xi_i}(q; t, \kappa)/\varphi_{\xi_j}(q; t, \kappa))}{\partial \xi_k} = 0. \quad (\text{A.67})$$

If $\varphi(q; t, \kappa) = \phi(U(q, t, \kappa))$ (as is always the case for CS, SG, and PE, and is the case when t is fixed for TTO and ES (see equations 2.3-2.15)), then (A.67) translates to

$$\frac{\partial(\phi_U U_{\xi_i}(q, t, \kappa)/\phi_U U_{\xi_j}(q, t, \kappa))}{\partial \xi_k} = 0 \quad (\text{A.68})$$

which reduces to

$$\frac{\partial(U_{\xi_i}(q, t, \kappa)/U_{\xi_j}(q, t, \kappa))}{\partial \xi_k} = 0 \quad (\text{A.69})$$

which is the condition for all ξ_k to be MPI in U . If U is homothetic in t , then $\varphi^i = \phi^i(\mu)$ (for TTO and ES). Then the same argument applies to μ as to U and the result is that all ξ_k are MPI in μ , which is equivalent to saying all ξ_k are MPI in U , since MPI is transformation independent.

Sufficiency: If utility satisfies MPI, then the QALY satisfies MPI.

The above argument holds in reverse since all steps are if and only if. \square

Proof of Lemma 5

Necessity: If the QALY satisfies SDI, then the utility function must satisfy SDI and the QALY must be an affine transform of the utility function.

If the QALY satisfies SDI, then

$$\varphi(q; t, \kappa) = \sum_j \varphi(\Xi_j; t, \kappa). \quad (\text{A.70})$$

Since

$$\varphi(q; t, \kappa) = \phi(U(q, t, \kappa)) \quad (\text{A.71})$$

$$\varphi(\Xi_j; t, \kappa) = \phi(U(\Xi_j, t, \kappa)) = \phi(v_j(\xi_j)) \quad (\text{A.72})$$

(from equations 2.3 to 2.15) (A.70) becomes

$$\phi(U(q, t, \kappa)) = \sum_{j=1}^K \phi(v_j(\xi_j)) \quad (\text{A.73})$$

or

$$U(q, t, \kappa) = f\left(\sum_{j=1}^K \phi(v_j(\xi_j))\right). \quad (\text{A.74})$$

Deriving the expressions for φ given in equations 2.3 to 2.15 using this functional form for utility, one obtains

$$\varphi(q; t, \kappa) = \phi\left(f\left(\sum v_j(\xi_j)\right)\right) \quad (\text{A.75})$$

$$\varphi(\Xi_j; t, \kappa) = \phi\left(f\left(v_j(\xi_j) + \sum_{k \neq j} v_k(\xi_k)\right)\right) = \phi\left(f(\tilde{v}_j(\xi_j))\right) \quad (\text{A.76})$$

such that (A.70) now becomes

$$\phi\left(f\left(\sum \tilde{v}_j(\xi_j)\right)\right) = \sum \left(\phi\left(f(\tilde{v}_j(\xi_j))\right)\right) \quad (\text{A.77})$$

which, given the definition of f , becomes

$$\phi\left(\sum \tilde{v}_j(\xi_j)\right) = \sum \left(\phi(\tilde{v}_j(\xi_j))\right) \quad (\text{A.78})$$

which is a Pexider equation with the solution $\phi(U) = aU + b$ (where $\phi(U) = \varphi$).

Sufficiency: If utility is affinely related to the QALY and exhibits SDI, then the QALY exhibits SDI.

If U satisfies SDI, then $U_{\xi_i, \xi_j}(q, t, \kappa) = 0$. Since $\varphi = aU + b$, this implies $\varphi_{\xi_i, \xi_j}(q; t, \kappa) = aU_{\xi_i, \xi_j}(q, t, \kappa) = 0$, which is the condition for φ to be SDI. \square

Appendix B

Simulation Results

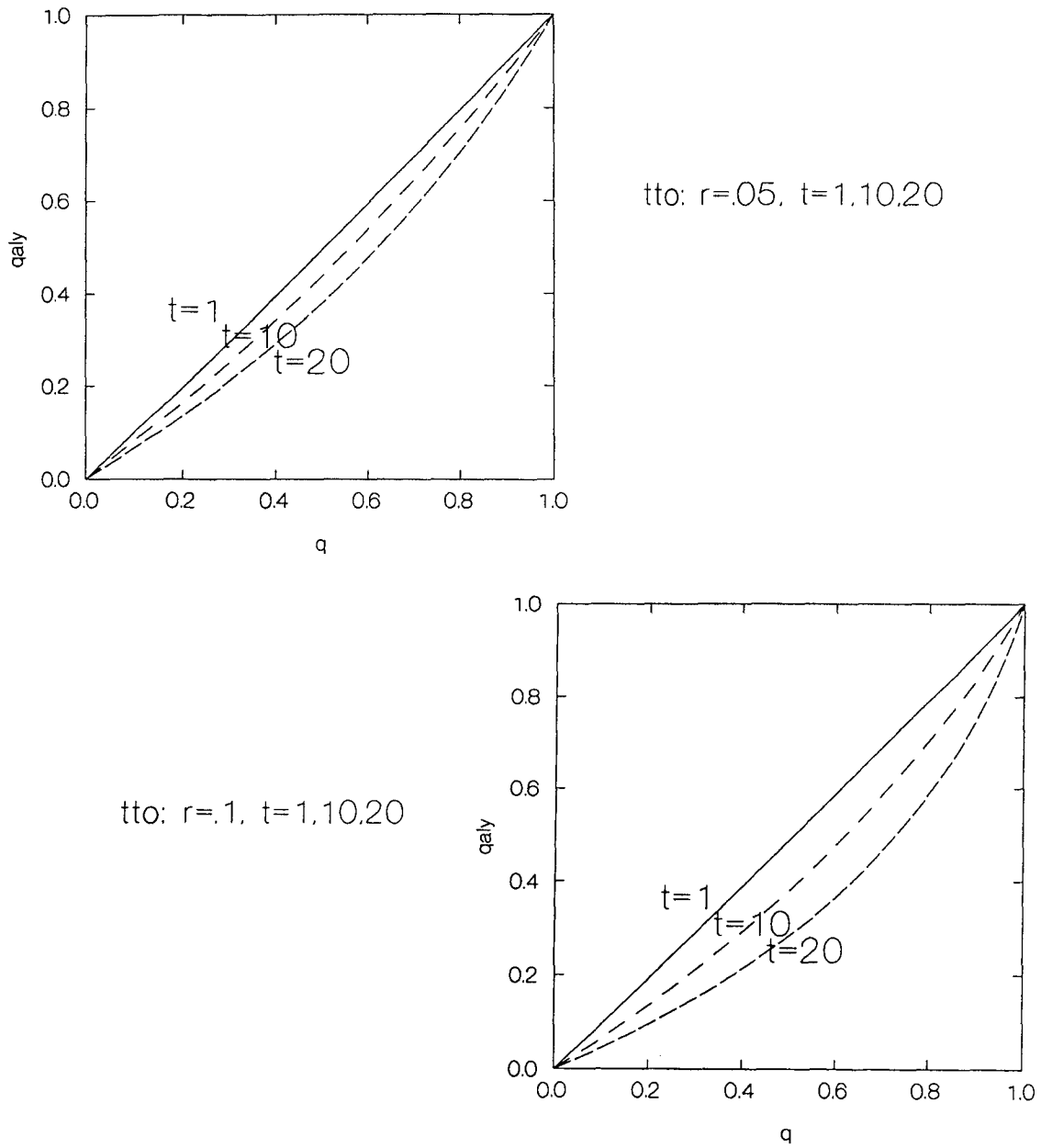


Figure B.1: Time Trade-Off

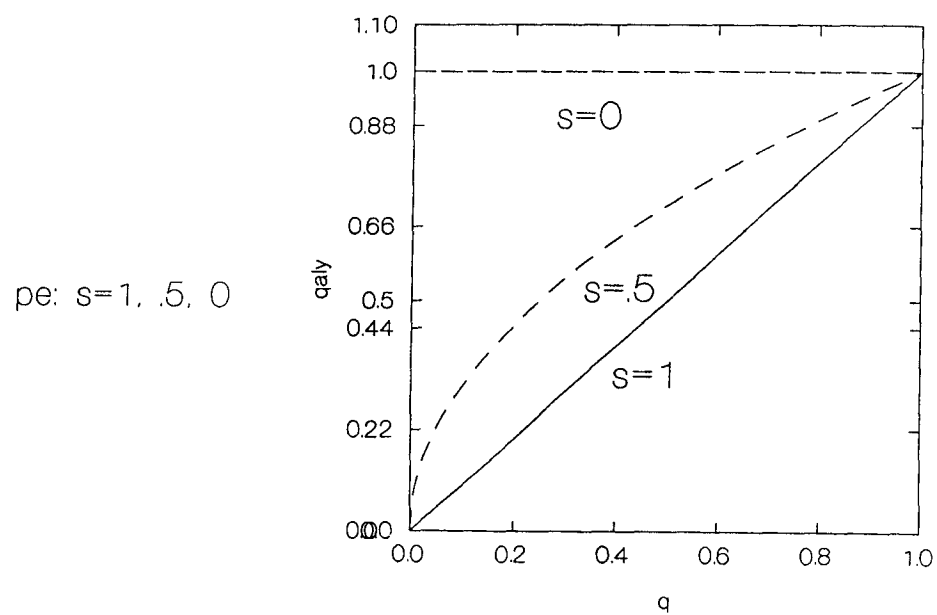
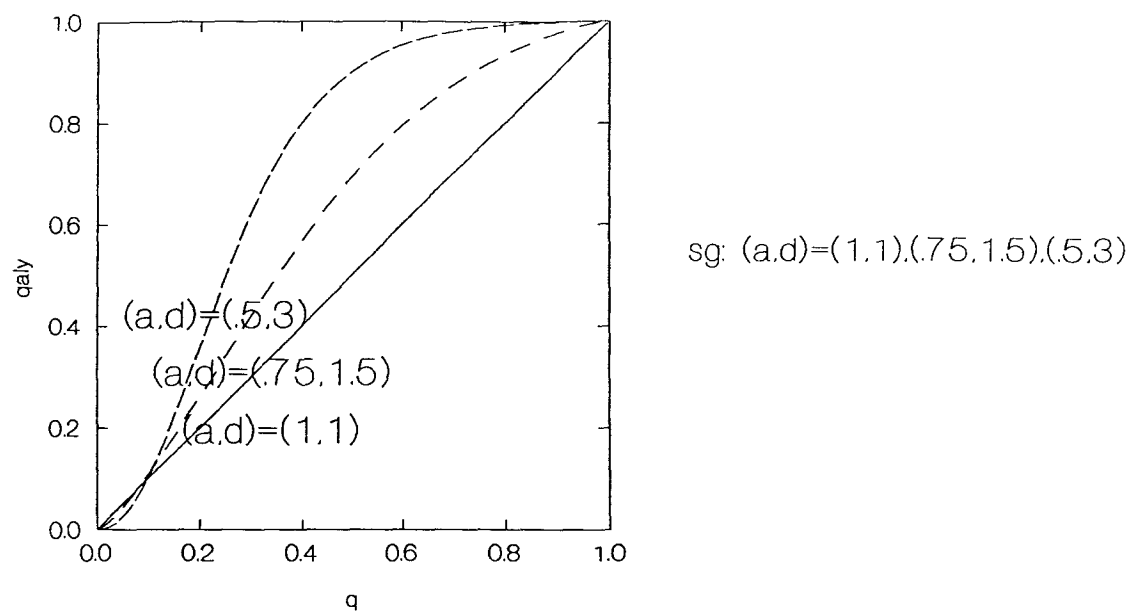


Figure B.2: Standard Gamble and Person Equivalents

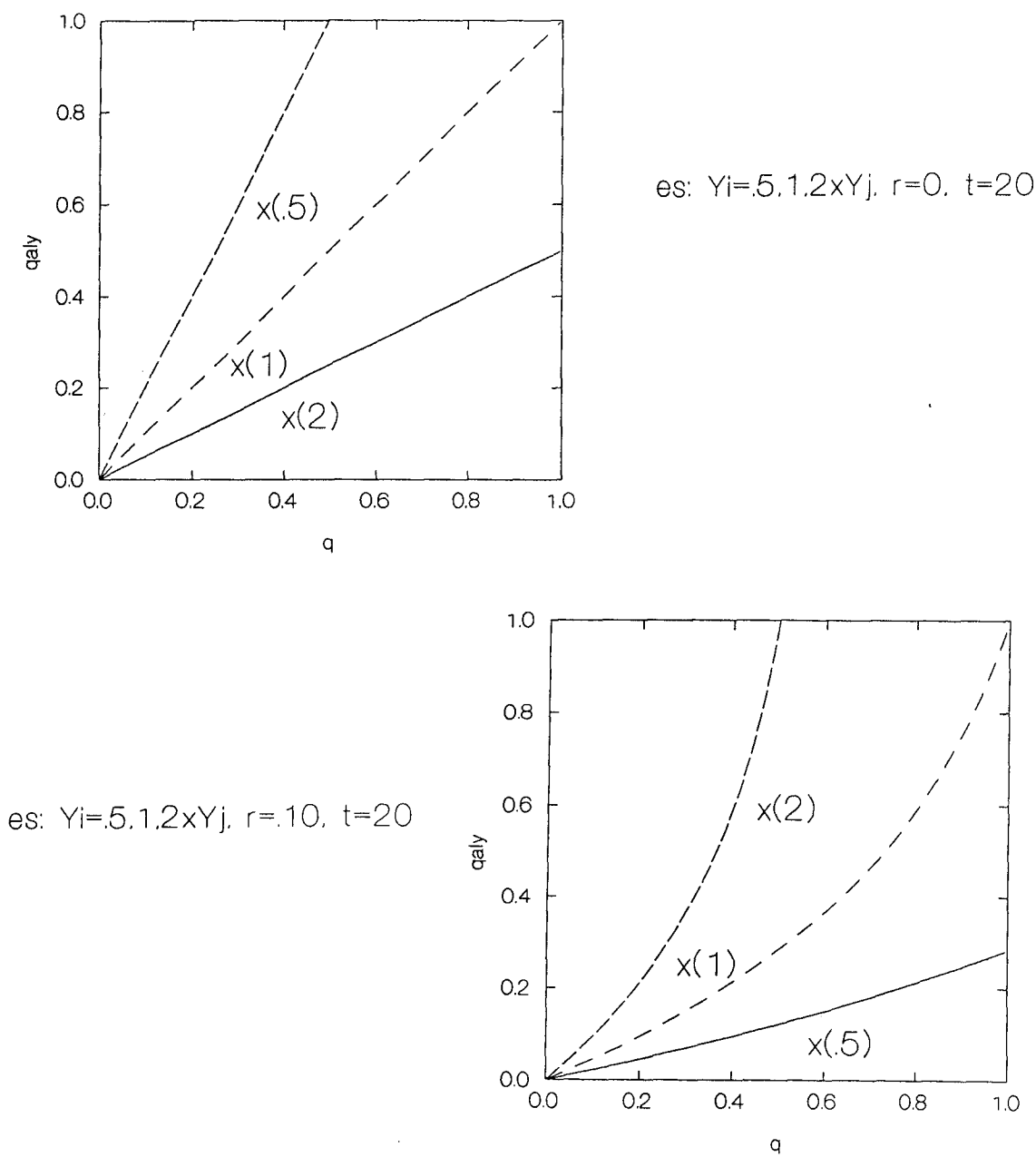


Figure B.3: Extended Sympathy

likely example

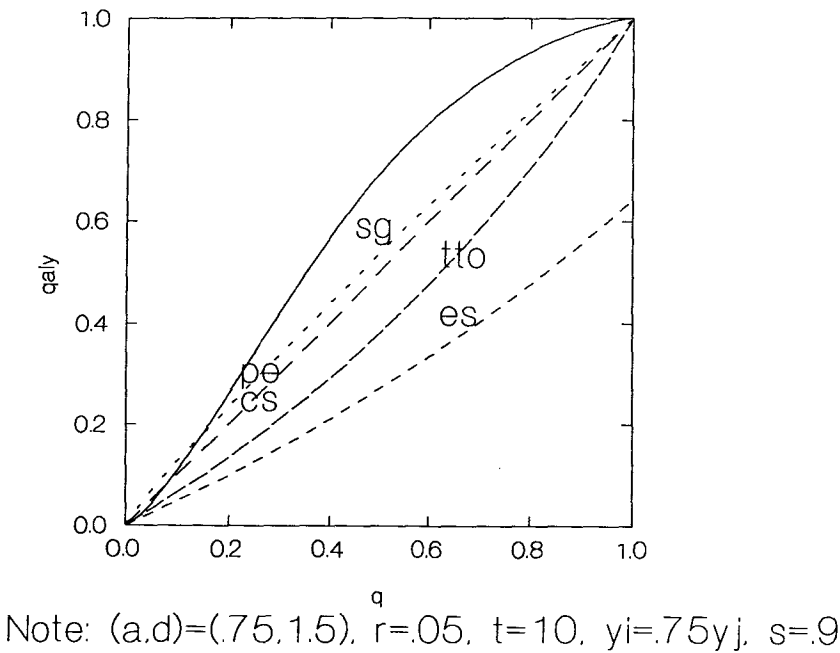


Figure B.4: Most Likely Case

Appendix C

Data Sources for Chapter 2

The data used in this paper are all drawn from the Canadian G.S.S. (General Social Survey) of 1985. This survey covered 11,200 respondents in a stratified sample of the Canadian population. The advantages of the scope of the study include a large number of observations from a cross-section of the population (as opposed to medical personnel or persons in a very localized area). Each individual is surveyed only for his or her own health state. This means (1) the observation advantage is much smaller than it might appear and (2) assumptions will have to be made that (like) individuals have the same preferences so that a utility function over several points in morbidity space can be estimated.

Variables

Utility, the dependent variable in this analysis, is taken to be the satisfaction with health variable (exists on a four-point scale). This is an ordered response variable with no interval properties (the categories were not assigned numerical ratings). Response rates on this variable were high (99 per cent), although the data are highly skewed towards higher levels of satisfaction (the smallest cell, very dissatisfied, had 500 observations).

The independent variables are various health characteristics. These form the arguments of the utility function. This assumes utility is based on levels of characteristics and does not depend on expectations or differences from some reference level (see Wright [1985] for supporting evidence on this stance). These variables fall into three

categories.

The first category is long-term activity limitation. These variables fit the very narrow definition of illness as physical limitation. There are four groups of variables pertaining to mobility (four), agility (three), sight (one), and hearing (one), as well as general activity limitation. These are multivariate binary variables: the ailment may be present or absent, and if present, moderate or severe. About 30 percent of the population reported some chronic ailment, so it should be expected that some cells will have few observations. While these four categories of ill-health are much more general than heretofore tested for in independence analysis, they are only a subset of those factors that affect satisfaction with health. For analysis at this level to be valid, one has to make assumptions about the distribution of these excluded variables.

The second category of independent variables consists of short term ailments, measured by their impact on activities *normally* undertaken in a two week period (hence, these measures should be independent of the long term measures above). While the "loss days" provide a numerical measure of severity of illness, these variables are less appealing than those in the first category because (1) the loss function depends on other non-health characteristics and reflects need as well as health itself, and (2) days may be lost to varying degrees which cannot be numerically compared (bed days are worse than restricted days, but it is not known by how much). Only 6 percent of respondents reported any bed days, while only 7 percent reported any restricted days. While few cells are well represented, it is believed the data are sufficient to determine if short run ailments are separable from more permanent ones. This does not appear to have been tested for before.

The final possible category of variables are more social in nature, including such factors as number of visits and contacts with friends and relatives. These are consistent with the broader WHO definition of health and Torrance's classification scheme

(1986). Unfortunately, emotional health is still absent. What needs to be determined is (1) whether these factors influence satisfaction over health, (2) if they do, is it in a fashion that would allow physical dysfunction to be analyzed in isolation, and (3) can these factors be combined in a piecemeal fashion with physical factors in broadly based QALY analysis?

Finally, age is used as a conditioning variable.

Construction of Variables

To estimate

$$U = \alpha + \sum \beta_l q_l^{0.5} + 0.5 \sum \sum \beta_{lk} q_l^{0.5} q_k^{0.5}$$

the variables are defined:

dependent variable

satisfaction with health

construction: survey item 73(a)

structure: ordered categorical (1=very satisfied, 2=satisfied, 3=dissatisfied, 4=very dissatisfied)

response: 11088/11200 (61 non-respondents in this category were also non-respondents in the independent variable category)

independent variables

chronic morbidity

endurance (E)

construction: poor health indicated (assigned a value of one) if poor health reported in any one of its factor components, otherwise good health indicated (assigned a value of zero), where the factor components are:

survey item q(27) 400m walk

survey item q(28) stair climb

survey item q(29) 5 kg carry

survey item q(30) standing

structure: binary

agility (A)

construction: poor health indicated (assigned a value of one) if poor health reported in any one of its factor components, otherwise good health indicated (assigned a value of zero), where the factor components are:

survey item q(31) bending

survey item q(33) grasping

survey item q(34) reaching

structure: binary

perception (P)

construction: poor health indicated (assigned a value of one) if poor health reported in any one of its factor components, otherwise good health indicated (assigned a value of zero), where the factor components are:

survey item q(35) seeing

survey item q(36) hearing

structure: binary

short term morbidity (S)

construction: poor health indicated (assigned a value of one) if any number of sick or bed days reported, otherwise good health indicated (assigned a value of zero). Factor components are:

survey item q(13) bed days

survey item q(17) activity limit days

structure: binary

social health (L)

construction: poor health indicated (assigned a value of one) if the respondent had fewer than average contacts or visits (average being the mean number of visits or contacts in the sample) across a majority of contact groups (eg. parents, children, siblings, friends, others); good health indicated (assigned a value of zero) if more than the average across most contact/visit groups. Factor components are:

visits (mother q(107), father (112), children (116), siblings (121),
other relatives (124), friends (127))

contacts (mother q(108), father (113), children (117), siblings (122),
other relatives (125), friends (128))

structure: binary

response rate 11064/11200

conditioning factors

age

construction: survey item 44

structure: ordered categorical (5 year age groups, 1=15-19 years, ,
14=80 years and over)

response: 11200/11200

Appendix D

Empirical Results for Chapter 2

Table D.1: Estimation Results

variable	all data	< 30yrs	30-65 yrs	> 65 yrs
constant	-.12061	.033990	-.08887	-.47021
(t-test)	(-7.5760)	(1.2451)	(-4.0536)	(-11.681)
endurance (E)	.55324	.70668	.57805	.70867
(t-test)	(13.773)	(7.0033)	(9.4992)	(10.297)
agility (A)	.31733	.56996	.33122	.36632
(t-test)	(6.001)	(4.4172)	(4.2674)	(3.8977)
perception (P)	.16695	.29324	.31818	.25607
(t-test)	(3.5098)	(2.2423)	(4.1021)	(3.4210)
short (S)	.44632	.29584	.49361	.61696
(t-test)	(10.096)	(4.2780)	(7.2285)	(5.1158)
social (L)	.054817	.048678	.021129	-.079966
(t-test)	(1.2846)	(.61175)	(.37935)	(-.51763)
ExA	.15977	-.29063	.22885	.10513
(t-test)	(2.4791)	(-1.5221)	(2.2870)	(1.0070)
ExP	-.10795	-.034529	-.15608	-.18840
(t-test)	(-1.5324)	(-.15539)	(-1.2294)	(-1.8560)
ExS	.23008	-.11095	.25493	.21252
(t-test)	(3.0861)	(-.64301)	(2.2306)	(1.4803)
ExL	.051348	-.16422	.14384	.050799
(t-test)	(.52604)	(-.67486)	(1.1037)	(.23410)
AxP	.033642	-.20137	-.014383	.082248
(t-test)	(.47541)	(-.77462)	(-.10467)	(.83324)
AxS	.11732	.20946	.049534	-.016288
(t-test)	(1.5140)	(1.0354)	(.41348)	(-.12960)
AxL	-.14345	-.21466	-.24133	.20035
(t-test)	(-1.3012)	(-.82644)	(-1.5902)	(.87907)
PxS	-.090396	.12764	-.041435	-.23521
(t-test)	(-1.1311)	(.54995)	(-.28075)	(-2.0565)

PxL (t-test)	.28658 (2.4075)	.58248 (2.0517)	.34481 (1.7705)	.087443 (.42160)
SxL (t-test)	-.15281 (-1.3523)	.0095979 (.043895)	-.17761 (-1.1314)	-.35416 (-1.2877)
a_1 (t-test)	1.4828 (81.057)	1.6859 (43.354)	1.5164 (54.981)	1.2962 (39.667)
a_2 (t-test)	2.3947 (79.262)	2.7675 (33.047)	2.4709 (52.248)	2.1416 (45.104)
likelihood	-10693	-2624.7	-4814.3	-3131.1
d.o.f.	11064	2913	5062	3089

Tests

Table D.2: Additive Tests

-	all data	< 30 yrs	30-65 yrs	> 65 yrs
ExA	2.48*	-1.52	2.29*	1.01
ExP	-1.53	-.155	-1.23	-1.86*
ExS	3.09*	-.643	2.23*	1.48
ExL	.526	-.675	1.10	.234
AxP	.475	-.775	-.105	.833
AxS	1.51	1.04	.413	-.130
AxL	-1.30	-.836	-1.59	.879
PxS	-1.13	.550	-.281	-2.06*
PxL	2.41*	2.05*	1.77*	.422
SxL	-1.35	.044	-1.13	-1.29

Note: t-test, degrees of freedom 11064, 2895, 5046, and 3071. “*” indicates significance at 5 per cent level.

Table D.3: Joint Additive Tests

all characteristics (10 dof)	43.16*	9.82	24.87*	14.23
chronic-short-social (6 dof)	27.56*	6.83	13.33*	7.74
physical-social (4 dof)	8.99	5.17	7.80	2.56
chronic-social (3 dof)	7.06	5.17	6.21	1.61
chronic-short (3 dof)	21.38*	1.50	7.75*	5.10
within chronic (3 dof)	7.78	3.38	6.51	4.09

Note: Wald test

Note: chronic= {E,A,P}, short= {S}, physical= {E,A,P,S}, social= {L}.

Table D.4: Multiplicative Tests (all data)

-	ExP	ExS	ExL	AxP	AxS	AxL	PxS	PxL	SxL
ExA	4.62*	.001	.049	.033	.012	3.03	3.37	4.31*	2.16
ExP		4.88*	.760	1.06	3.76*	1.20	.001	6.08*	1.13
ExS			.047	1.17	.016	1.99	3.08	4.84*	2.23
ExL				.076	.058	1.52	.752	4.31*	1.85
AxP					.016	1.74	1.07	4.51*	1.78
AxS						1.96	2.38	4.30*	2.11
AxL							1.13	5.83*	.058
PxS								5.94*	1.03
PxL									6.81*

Note: Wald test, 1 degree of freedom.

Table D.5: Joint Tests

chronic-short	3.09	(3 dof)
chronic-social	6.57	(3 dof)
physical-social	5.03	(6 dof)
within chronic	(sing. matrix)	

Note: Wald test.

Note: joint tests where the number of individual conditions exceeds the number of estimated parameters cannot be done.

Table D.6: Multiplicative Tests (< 30 yrs)

-	ExP	ExS	ExL	AxP	AxS	AxL	PxS	PxL	SxL
ExA	.254	.044	.344	.092	2.06	.120	.643	3.24	.009
ExP		.084	.407	.263	.525	.612	.295	3.82*	.003
ExS			.368	.151	1.14	.601	.476	3.17	.006
ExL				.248	.581	.054	.554	3.99*	.106
AxP					1.36	.454	.699	3.49	.016
AxS						.801	.005	2.48	.001
AxL							.737	3.75*	.196
PxS								2.96	.002
PxL									1.92

Note: Wald test, 1 degree of freedom.

Table D.7: Joint Tests

chronic-short	1.30	(3 dof)
chronic-social	4.77	(3 dof)
physical-social	5.04	(4 dof)
within chronic	(sing. matrix)	

Note: Wald test.

Table D.8: Multiplicative Tests (30-65 yrs)

-	ExP	ExS	ExL	AxP	AxS	AxL	PxS	PxL	SxL
ExA	4.45*	.119	.711	.749	.763	5.44*	1.64	2.41	1.26
ExP		4.01*	1.37	.187	1.37	2.21	.217	3.18	1.11
ExS			.863	.981	.306	2.49	1.10	2.68	1.31
ExL				1.03	.946	2.86	1.17	1.40	2.18
AxP					.086	2.24	.005	2.71	1.16
AxS						2.50	.201	2.51	1.26
AxL							2.17	4.70*	.387
PxS								2.98	1.15
PxL									3.81*

Note: Wald test, 1 degree of freedom.

Table D.9: Joint Tests

chronic-short	1.22	(3 dof)
chronic-social	5.41	(3 dof)
physical-social	(sing. matrix)	
within chronic	(sing. matrix)	

Note: Wald test.

Table D.10: Multiplicative Tests (> 65 yrs)

-	ExP	ExS	ExL	AxP	AxS	AxL	PxS	PxL	SxL
ExA	4.39*	.016	.094	.151	.407	.486	5.97*	.199	1.25
ExP		4.78*	.001	2.07	1.81	.514	.238	.112	1.67
ExS			.102	.043	.520	.811	5.56*	.205	1.35
ExL				.132	.046	.350	.025	.096	1.26
AxP					.599	.812	3.23	.221	.872
AxS						.733	1.91	.169	1.53
AxL							.409	.031	1.84
PxS								.085	1.79
PxL									.934

Note: Wald test, 1 degree of freedom.

Table D.11: Joint Tests

chronic-short	5.71	(3 dof)
chronic-social	.415	(3 dof)
physical-social	2.31	(3 dof)
within chronic	(sing. matrix)	

Note: Wald test.

Table D.12: Multilinear Tests

$\frac{\partial MRS}{\partial q}$	all data	< 30 yrs	30-65 yrs	> 65 yrs
(E,A),P	.217	.734	.155	.000
(E,A),S	2.27	.582	2.53	3.32
(E,A),L	5.11+	.294	6.20+	.000
(E,P),A	3.36	.842	4.07+	1.38
(E,P),S	9.52+	.000	3.73	10.04+
(E,P),L	.981	1.27	.293	.184
(E,S),A	1.16	.495	2.30	1.19
(E,S),P	.720	.014	.010	.656
(E,S),L	3.94+	.155	3.56	1.55
(E,L),A	2.01	.346	.183	.556
(E,L),P	.866	.274	.120	.545
(E,L),S	2.23	.016	.196	.098
(A,P),E	15.15+	.015	6.09+	9.85+
(A,P),S	4.72+	.241	.235	1.84
(A,P),L	2.97	1.62	5.13+	.211
(A,S),E	3.61	.302	.915	7.11+
(A,S),P	.082	.112	.581	.021
(A,S),L	1.15	.006	1.86	.213
(A,L),E	.037	.314	.064	.721
(A,L),P	.775	.285	.117	.945
(A,L),S	1.97	.056	.182	.109
(P,S),E	44.87+	.073	13.59+	31.96+
(P,S),A	12.65+	1.75	1.98	2.74
(P,S),L	.220	2.52	1.46	.191
(P,L),E	.221	.343	.098	.223
(P,L),A	1.43	.273	.169	.727
(P,L),S	.915	.054	.172	.161
(L,S),E	.050	.316	.068	.729
(L,S),A	1.91	.489	.174	.512
(L,S),P	.882	.260	.117	.492

Table D.13: Joint Tests

chronic-short (3 dof)	11.16+	.4.26	5.26	12.26+
chronic-social (3 dof)	10.41+	2.20	7.25	2.20
physical-social (6 dof)	17.08+	10.25	17.47+	140.86+
w/in chronic (3 dof)	27.72+	1.31	12.35+	13.82+

Note: Wald test. One degree of freedom for single tests. “+” indicates test is significantly different from zero, but sign is consistent with WDI.

Simulated QALY Values

Table D.14: Simulated Holistic QALY Values

q	cs	sg	tto	es	pe
\bar{q}	1.000	1.000	1.000	0.643	1.000
E	0.759	0.912	0.653	0.446	0.780
A	0.868	0.965	0.796	0.530	0.880
P	0.933	0.987	0.891	0.584	0.939
S	0.809	0.939	0.717	0.484	0.827
L	0.979	0.997	0.964	0.623	0.981
ExA	0.514	0.708	0.393	0.279	0.549
ExP	0.733	0.895	0.619	0.425	0.753
ExS	0.404	0.572	0.295	0.212	0.442
ExL	0.707	0.879	0.592	0.408	0.732
AxP	0.776	0.922	0.674	0.458	0.796
AxS	0.594	0.789	0.470	0.330	0.625
AxL	0.907	0.979	0.851	0.562	0.916
PxS	0.773	0.920	0.671	0.457	0.793
PxL	0.780	0.924	0.680	0.462	0.800
SxL	0.854	0.960	0.776	0.519	0.868

aggregated values

Table D.15: Aggregated Values (cs)

q	holistic/multilinear	additive	multiplicative
ExA	0.514	0.627	0.564
ExP	0.730	0.692	0.660
ExS	0.404	0.568	0.479
ExL	0.707	0.737	0.727
AxP	0.776	0.801	0.784
AxS	0.594	0.677	0.628
AxL	0.907	0.846	0.841
PxS	0.773	0.742	0.717
PxL	0.780	0.911	0.909
SxL	0.854	0.788	0.780

Note: $\lambda = 1.944$.

Table D.16: Aggregated Values (sg)

q	holistic/multilinear	additive	multiplicative
ExA	0.708	0.877	0.805
ExP	0.895	0.899	0.872
ExS	0.572	0.851	0.726
ExL	0.879	0.909	0.904
AxP	0.922	0.952	0.941
AxS	0.789	0.904	0.855
AxL	0.979	0.962	0.960
PxS	0.920	0.926	0.907
PxL	0.924	0.984	0.983
SxL	0.960	0.937	0.933

Note: $\lambda = 23.415$.

Table D.17: Aggregated Values (tto)

q	holistic/multilinear	additive	multiplicative
ExA	0.393	0.449	0.445
ExP	0.619	0.544	0.542
ExS	0.295	0.370	0.364
ExL	0.592	0.617	0.616
AxP	0.674	0.686	0.685
AxS	0.470	0.512	0.509
AxL	0.851	0.759	0.759
PxS	0.671	0.607	0.606
PxL	0.680	0.854	0.854
SxL	0.776	0.680	0.679

Note: $\lambda = 0.05469$.

Table D.18: Aggregated Values (es)

q	holistic/multilinear	additive	multiplicative
ExA	0.279	0.333	0.259
ExP	0.425	0.387	0.348
ExS	0.212	0.287	0.182
ExL	0.408	0.427	0.414
AxP	0.458	0.471	0.449
AxS	0.330	0.372	0.312
AxL	0.562	0.511	0.504
PxS	0.457	0.425	0.394
PxL	0.462	0.565	0.561
SxL	0.519	0.465	0.455

Note: $\lambda = 3.361$.

Table D.19: Aggregated Values (pe)

q	holistic/multilinear	additive	multiplicative
ExA	0.549	0.660	0.590
ExP	0.753	0.719	0.684
ExS	0.442	0.607	0.505
ExL	0.732	0.761	0.749
AxP	0.796	0.820	0.800
AxS	0.625	0.707	0.652
AxL	0.916	0.861	0.855
PxS	0.793	0.766	0.738
PxL	0.800	0.920	0.917
SxL	0.868	0.807	0.798

Note: $\lambda = 2.661$.

Measures of distortion

Note: The measures of distortion do not take into account whether the calculated differences are statistically significant or not. The distortions are measured from the holistic value for the instrument indicated at the top over all fifteen morbid states.

Table D.20: Distortions from cs Values

aggregation	cs	sg	tto	es	pe
holistic	0.000	1.962	1.345	4.962	0.296
additive	0.744	2.618	1.436	4.833	0.871
multiplicative	0.583	2.302	1.455	5.198	0.689

Table D.21: Distortions from sg Values

aggregation	cs	sg	tto	es	pe
holistic	1.962	0.000	3.307	6.925	1.667
additive	1.611	.735	3.250	6.795	1.382
multiplicative	1.911	.503	3.269	7.160	1.652

Table D.22: Distortions from tto Values

aggregation	cs	sg	tto	es	pe
holistic	1.345	3.307	0.000	3.618	1.640
additive	1.706	3.962	.712	3.488	1.994
multiplicative	1.416	3.647	.702	3.854	1.655

Table D.23: Distortions from es Values

aggregation	cs	sg	tto	es	pe
holistic	4.962	6.925	3.618	0.000	5.258
additive	5.314	7.580	3.675	.479	5.612
multiplicative	5.013	7.265	3.655	.446	5.272

Table D.24: Distortions from pe Values

aggregation	cs	sg	tto	es	pe
holistic	.296	1.667	1.640	5.258	0.000
additive	.699	2.322	1.692	.5128	.707
multiplicative	.572	2.007	1.711	5.494	.525

Appendix E

Proofs to Chapter 3

Proof Lemma 1 (HK): To be exact, $w(q)L^*(u, p, w(q), q)$ cannot depend on the variable arguments of U (this would make HK a non-monotonic transform of U). Thus, q must be the same for all states compared, i.e. $V(p, w(q), w(q)T + I, q) = \dot{V}(p, w, wT + I)$. Then $u^A \geq u^B \leftrightarrow \dot{V}(p, w, wT^A + I) \geq \dot{V}(p, w, wT^B + I) \leftrightarrow wT^A + I \geq wT^B + I$. By duality relationships: $L^*(U, p, w) = \dot{L}^*(\bar{V}(p, w, wT + I), p, w) = \bar{L}(wT + I, p, w)$. Then exactness requires $\bar{L}(wT^A + I, p, w) \geq \bar{L}(wT^B + I, p, w) \leftrightarrow u^A \geq u^B$. This requires that $\partial \bar{L} / \partial I = \partial L^* / \partial I > 0$ (i.e. labour is a normal good). This result is unchanged by the presence of leisure constraints (as can be seen if the leisure constrained indirect utility function is substituted for the unconstrained indirect utility function above. However, exactness does not hold if the set of binding constraints changes because this makes HK a function of the variable arguments of U .

In the case where labour is constrained, the proof proceeds as above, except τ_L replaces T . No restriction on income elasticities is required: the transform is positive since, *ceteris paribus*, utility increases with the constraint because time can be better allocated. \square

Proof Lemma 1 (WTP): WTP may be expressed:

$$\mathcal{V}(p, q^A, T^A, I - cv) = u^B. \quad (\text{E.1})$$

To be exact, cv must be positive whenever state A is preferred to state B. Let $u^A > u^B$. Then for (E.1) to hold, income must be changed in state A to reduce the utility of

state A. By assumption, $\partial \mathcal{V} / \partial I > 0$, so this requires that income be reduced, or alternately, $cv > 0$. Thus, this is exact. \square

Proof Lemma 1 (HYE): HYE may be expressed:

$$\Delta HYE = (T^A - m^A) - (T^B - m^B), \text{ where}$$

$$\mathcal{V}(p, \bar{q}, T^i - m^i, I) = u^i. \quad (\text{E.2})$$

To be exact, $T^i - m^i$ must be larger the larger is u^i . Let u^i increase by du^i . Then the equality in (E.2) requires that $T^i - m^i$ be changed to increase the left hand side of the equation by du^i . Since \mathcal{V} is assumed to be increasing in T (not unreasonable at a reference morbid state of perfect health), this requires $T^i - m^i$ increase. Thus, this is exact.¹ \square

Proof Lemma 2: Consider a project affecting longevity alone (i.e. $(q, T^B) \rightarrow (q, T^A)$). The true willingness to pay in time units is given by $U(q, T^A - m) = U(q, T^B)$ which is, of course, $m = T^A - T^B$. The HYE approach instead sets $U(q, T^A) = U(\bar{q}, T^A - m^A)$, or $m^A = MRS_{q,T}(\bar{q}, T^A)dq$, and $U(q, T^B) = U(\bar{q}, T^A - m^B)$, or $m^B = MRS_{q,T}(\bar{q}, T^B)dq$, and the value of the health status change is $m^B - m^A$. Given that $m^B - m^A = -\frac{\partial MRS_{q,T}}{\partial T}dq = -(\frac{V_{qT}}{V_T} + \frac{V_q}{V_T}R_{TT}^A)dq$, the HYE value is falling as the absolute risk aversion with respect to time increases.

Now consider a project affecting q alone (i.e. $(q^B, T) \rightarrow (q^A, T)$). The true willingness to pay in time units is given by $U(q^A, T - m) = U(q^B, T)$. The HYE approach instead sets $U(q^A, T) = U(\bar{q}, T - m^A)$, or $m^A = MRS_{q,T}(\bar{q}, T)dq^A$, and $U(q^B, T) = U(\bar{q}, T - m^B)$, or $m^B = MRS_{q,T}(\bar{q}, T)dq^B$, and the value of the health status change is $m^B - m^A$. Given that $m^B - m^A = -MRS_{q,T}(\bar{q}, T)(\bar{q} - q^A - \bar{q} + q^B)$ and that $m = MRS_{q,T}(q^A, T)(q^A - q^B)$, $m > / < m^B - m^A$ as $\frac{\partial MRS_{q,T}}{\partial q} < / > 0$. But

¹A problem may arise if the wage schedule varies with time. In this case, the marginal value of time varies, and as a result, HYE can be based on two different transformations of the utility function where the transformation depends on the level of utility (i.e. is inexact).

since $\frac{\partial MRS_{qT}}{\partial q} = \frac{V_T}{V_q}(-R_{qq}^A) - \frac{V_{qT}}{V_q}$, this implies the HYE value is falling as the absolute risk aversion with respect to morbidity increases. \square

Proof to Corollary 1: If $\sum_i HK_i(T_i)$ is consistent with a Bergson-Samuelson social welfare function, then

$$\begin{aligned} \sum_i w_i L_i(v_i(T_i^A), x_i, y) &\geq \sum_i w_i L_i(v_i(T_i^B), x_i, y) \leftrightarrow \\ \sum_i w'_i L_i(v_i(T_i^A), x'_i, y') &\geq \sum_i w'_i L_i(v_i(T_i^B), x'_i, y') \end{aligned} \quad (E.3)$$

where x_i are person-specific variables, y are variables held in common by all persons (prime denotes a change in value), and $v(T) = V(p, w, wT + I, q)$.

For the wage rate, this requires that $w_i(\partial L_i / \partial w) + L_i = 0$ for all i . For other elements of (x_i, y) , (3.20) requires $L_i(v(T_i), x_i, y) = b(y)(l_i(v_i(T_i)) + a_i(x_i, y))$. But this requires that $\frac{\partial L_i(I_i)}{\partial T_i} = \frac{\partial L_i(I'_i)}{\partial T_i}$. Yet, when income rises, the level of consumption of purchased goods goes up and, normally the marginal value of leisure increases relative to the marginal value of purchased goods (so more time is allocated to leisure and less to labour). This does not happen only if leisure has no utility, which implies $L^* = T$ (the case where the marginal utility of purchased goods is zero can be ruled out since $\partial L / \partial T > 0$ by assumption). But if $L^* = T$, $\partial L / \partial w = 0$ and the wage constraint holds if and only if $dw = 0$ ($w0 + dwT = 0$) and $w_i = w$ for all i .

In the labour constrained case, $u_i = \bar{v}_i(\tau_L(q_i))$, so $\tau_L(q_i) = \phi^i(u_i)$. Then the human capital statistic becomes $HK_i(q_i) = w_i \tau_L(q_i) = w_i f^i(u_i, x_i, y)$. Then (3.20) requires that $w_i f^i(u_i, x_i, y) = a(y)\gamma^i(u_i) + b_i(x_i, y)$. But this implies (1) $w_i = w$ for all i , and (2) $u_i \stackrel{o}{=} \frac{w\tau_L(q_i) - b_i(x_i, y)}{a(y)}$. This means that $T - \tau_L(q_i)$ does not enter utility (i.e. leisure has no utility; if it did, this representation of preferences implies that utility would be decreasing in the constraint). \square

Proof of Corollary 2: From (3.20), $cv(q_i, T_i) = a(y)\gamma^i(u_i) + b_i(x_i, y)$. For the move from any health state to itself, this may be re-expressed as $u_i \stackrel{o}{=} a(y)I_i + b_i(x_i, y)$,

where x_i includes (q_i, T_i) , but not I_i . Note that this applies to the global indirect utility function, \mathcal{V} , and thus applies to all constrained and unconstrained cases. \square

Proof of Corollary 3: From (3.20), $HYE(q_i, T_i) = a(y)\gamma^i(u_i) + b_i(x_i, y)$. At \bar{q} (the reference health state), this may be re-expressed as $u_i \stackrel{o}{=} a(y)T_i + b_i(x_i, y)$. x_i includes (I_i) , but not (q_i, T_i) . The argument for excluding q_i is based on the assumption that all morbid states are indifferent in death because they are not actually experienced. Hence, if $T_i = 0$, then no utility from q_i can be realized (bequests allow utility to be derived from wealth even after death, however). The restriction derived from (3.20) then seems to imply that q_i be the same for all individuals (i.e. an element of y). But this restriction was derived when q_i was set equal to the reference health state, which is fixed for all evaluations, and only applies to this reference health state. The marginal utility of time spent in the reference health state must be the same for all individuals (i.e. a year of perfect health is worth the same to everyone). Theoretically, any q may be chosen as the reference health level, so this restriction must hold for any q . \square

Appendix F

Proofs to Chapter 4

Theorem 1:

Necessity: If \mathcal{R}_Γ is independent of the reference point, \bar{q} , then

$$\begin{aligned} & (Q^A, T^A) \mathcal{R}_\Gamma (Q^B, T^B) \\ & \leftrightarrow \Gamma(f_1(U^1(q_1^A, t_1^A), \bar{q}), \dots, f_N(U^N(q_N^A, t_N^A), \bar{q})) \geq \\ & \Gamma(f_1(U^1(q_1^B, t_1^B), \bar{q}), \dots, f_N(U^N(q_N^B, t_N^B), \bar{q})) \end{aligned} \quad (\text{F.1})$$

for all possible reference points, \bar{q} . If \mathcal{R}_Γ is independent of \bar{q} , then an arbitrary fixed reference point, \hat{q} , may be chosen from the set of reference points. Let $f_i(U^i(q_i, t_i), \hat{q}) = g_i(u_i)$. Substitute this into (F.1) to obtain

$$\Gamma(g_1(u_1^A), \dots, g_N(u_N^A)) \geq \Gamma(g_1(u_1^B), \dots, g_N(u_N^B)). \quad (\text{F.2})$$

Then there exists some function, $W(u_1, \dots, u_N) := \Gamma(g_1(u_1), \dots, g_N(u_N))$, such that

$$(Q^A, T^A) \mathcal{R}_\Gamma (Q^B, T^B) \leftrightarrow W(u_1^A, \dots, u_N^A) \geq W(u_1^B, \dots, u_N^B). \quad (\text{F.3})$$

Sufficiency: If \mathcal{R}_Γ is consistent with a Bergson-Samuelson social welfare function, it must be independent of the reference point, i.e.

$$(Q^A, T^A) \mathcal{R}_\Gamma (Q^B, T^B) \leftrightarrow W(u_1^A, \dots, u_N^A) \geq W(u_1^B, \dots, u_N^B). \quad (\text{F.4})$$

□

Theorem 2: Additively separable case.

From the proof of Theorem 1, it is possible to obtain

$$\begin{aligned}
& (Q^A, T^A) \mathcal{R}_\Gamma (Q^B, T^B) \\
& \leftrightarrow W(u_1^A, \dots, u_N^A) \geq W(u_1^B, \dots, u_N^B) \\
& \leftrightarrow \sum_i \phi^i(g_i(u_i^A)) \geq \sum_i \phi^i(g_i(u_i^B)) \\
& \leftrightarrow \sum_i \phi^i(f_i(u_i^A, \bar{q})) = \sum_i \phi^i(f_i(u_i^B, \bar{q}))
\end{aligned} \tag{F.5}$$

for all \bar{q} . This means that

$$\sum_i \phi^i(f_i(u_i, \bar{q})) = \Phi(\sum_i \phi^i(g_i(u_i)), \bar{q}), \tag{F.6}$$

where Φ is increasing in its first argument.

Define $z_i = \phi^i(g_i(u_i))$, and $h_i(z_i, \bar{q}) = \phi^i(f_i(u_i, \bar{q}))$ where z_i is a continuous variable and h_i is increasing. Substitute this into (F.6),

$$\sum_i h_i(z_i, \bar{q}) = \Phi(\sum_i z_i, \bar{q}). \tag{F.7}$$

Because each h_i is continuous in z_i , Φ is continuous in its first argument and, for each \bar{q} , (F.7) is a Pexider equation with the solution

$$h_i(z_i, \bar{q}) = a(\bar{q})z_i + b_i(\bar{q}) \tag{F.8}$$

for all \bar{q} (see Eichhorn [1978], Theorem 3.1.5). This implies

$$\phi^i(f_i(u_i, \bar{q})) = a(\bar{q})\phi^i(g_i(u_i)) + b_i(\bar{q}) \tag{F.9}$$

or that

$$\phi^i(M^i(q_i, t_i, \bar{q})) = a(\bar{q})\phi^i(g_i(U^i(q_i, t_i))) + b_i(\bar{q}). \tag{F.10}$$

Set $q_i = \bar{q}$ in (F.10). By definition, $f_i(u_i, \bar{q}) = t_i$ at this value of q (i.e. no t is given up for the move from \bar{q} to \bar{q}). Thus (F.10) becomes

$$\phi^i(t_i) = \bar{a}(q_i)\phi^i(g_i(U^i(q_i, t_i))) + \bar{b}_i(q_i) \tag{F.11}$$

or

$$\phi^i(g_i(U^i(q_i, t_i))) = a(q_i)\phi^i(t_i) + b_i(q_i), \quad (\text{F.12})$$

where $a(q_i) = 1/\bar{a}(q_i)$ and $b_i(q_i) = -\bar{b}_i(q_i)/\bar{a}(q_i)$. It follows that

$$U^i(q_i, t_i) = \Phi^i(a(q_i)\phi^i(t_i) + b_i(q_i)), \quad (\text{F.13})$$

where Φ^i is increasing. This implies (4.13).

Sufficiency:

Suppose that (4.13) holds. Then

$$m_i = M^i(q_i, t_i, \bar{q}) = \theta^i \left[\frac{a(q_i)\phi^i(t_i) + b_i(q_i) - b_i(\bar{q})}{a(\bar{q})} \right] \quad (\text{F.14})$$

(where θ^i is the inverse of the function ϕ^i) and

$$\sum_i \phi^i(m_i) = \frac{\sum_i [a(q_i)\phi^i(t_i) + b_i(q_i)] - \sum_i b_i(\bar{q})}{a(\bar{q})} \quad (\text{F.15})$$

so that

$$(Q^A, T^A)\mathcal{R}_\Gamma(Q^B, T^B) \leftrightarrow \sum_i [a(q_i^A)\phi^i(t_i^A) + b_i(q_i^A)] \geq \sum_i [a(q_i^B)\phi^i(t_i^B) + b_i(q_i^B)] \quad (\text{F.16})$$

which makes \mathcal{R}_Γ independent of \bar{q} . \square

Theorem 3: Additive Case

From Theorem 2, independence for any additively separable case requires

$$U^i(q_i, t_i) \stackrel{\circ}{=} a(q_i)t_i + b_i(q_i). \quad (\text{F.17})$$

Set $t_i = 0$. Then Condition N requires that $b_i(q_i)$ be independent of q_i and (4.15) is immediate.

Sufficiency follows from Theorem 2. \square

Theorem 4: Cobb-Douglas case.

Combining the results of Theorem 1 with equation (4.16), if \mathcal{R}_Γ is independent of \bar{q} , then

$$\begin{aligned}
 & (Q^A, T^A) \mathcal{R}_\Gamma(Q^B, T^B) \\
 & \leftrightarrow \prod_i f_i(u_i^A, \bar{q}) \geq \prod_i f_i(u_i^B, \bar{q}) \\
 & \leftrightarrow \prod_i g_i(u_i^A) \geq \prod_i g_i(u_i^B) \\
 & \leftrightarrow \prod_i z_i^A \geq \prod_i z_i^B, \tag{F.18}
 \end{aligned}$$

where $g_i(u_i) = f_i(u_i, \hat{q})$ (\hat{q} being a fixed reference value of \bar{q}), and $z_i = g_i(u_i)$ for all $i = 1, \dots, N$. Defining $h_i(z_i, \bar{q}) = f_i(u_i, \bar{q})$, it follows that

$$\prod_i h_i(z_i, \bar{q}) = \phi(\prod_i z_i, \bar{q}). \tag{F.19}$$

For every \bar{q} , this is a Pexider equation with the solution

$$h_i(z_i, \bar{q}) = \bar{b}_i(\bar{q}) t_i^{\bar{a}(\bar{q})} \tag{F.20}$$

(see Eichhorn [1978], Theorem 3.5.5). Setting $q_i = \bar{q}$ as in Theorem 2 yields

$$t_i = \bar{b}_i(q_i) [g_i(u_i)]^{\bar{a}(q_i)}. \tag{F.21}$$

Rearranging,

$$u_i = U^i(q_i, t_i) \stackrel{\circ}{=} b_i(q_i) t_i^{a(q_i)}, \tag{F.22}$$

where

$$a(q_i) = 1/\bar{a}(q_i) \tag{F.23}$$

and

$$b_i(q_i) = \bar{b}(q_i)^{-1/\bar{a}(q_i)}. \tag{F.24}$$

Sufficiency is obvious from inspection. \square

Theorem 5:

If \mathcal{R}_Γ is independent of \bar{q} ,

$$\begin{aligned}
 & (Q^A, T^A)\mathcal{R}_\Gamma(Q^B, T^B) \\
 & \leftrightarrow \min\{m_1^A, \dots, m_N^A\} \geq \min\{m_1^B, \dots, m_N^B\} \\
 & \leftrightarrow \min\{f_1(u_1^A, \bar{q}), \dots, f_N(u_N^A, \bar{q})\} \geq \min\{f_1(u_1^B, \bar{q}), \dots, f_N(u_N^B, \bar{q})\} \\
 & \leftrightarrow \min\{g_1(u_1^A), \dots, g_N(u_N^A)\} \geq \min\{g_1(u_1^B), \dots, g_N(u_N^B)\} \\
 & \leftrightarrow \min\{z_1^A, \dots, z_N^A\} \geq \min\{z_1^B, \dots, z_N^B\}, \tag{F.25}
 \end{aligned}$$

where $g_i(u_i) = f_i(u_i, \hat{q})$, \hat{q} is some fixed reference value of \bar{q} , and $z_i = g_i(u_i)$ for all $i = 1, \dots, N$. Defining $h_i(z_i, \bar{q}) = f_i(u_i, \bar{q})$, it follows that

$$\min\{h_1(z_1, \bar{q}), \dots, h_N(z_N, \bar{q})\} = F(\min\{z_1, \dots, z_N\}, \bar{q}). \tag{F.26}$$

This requires

$$h_i(z_i, \bar{q}) = h_k(z_k, \bar{q}) \leftrightarrow z_i = z_k. \tag{F.27}$$

The range condition ensures that there are z 's satisfying (F.27) at every level. Therefore $h_i = h_k = h$ for all i, k , and

$$f_i(U^i(q_i, t_i), \bar{q}) = h_i(g_i(u_i, \bar{q})) = h(g_i(u_i), \bar{q}). \tag{F.28}$$

Setting $\bar{q} = q_i$ as in Theorem 2, the following is obtained

$$t_i = h(g_i(u_i, q_i)) \tag{F.29}$$

and, therefore

$$u_i = g_i^{-1}(v(q_i, t_i)) \triangleq v(q_i, t_i) \tag{F.30}$$

for some function v . This implies (4.19).

Sufficiency is immediate. \square

Theorem 6 (Blackorby and Donaldson [1988]): Concavity

Recall that m is defined by the following

$$m = \min_t \{t/U(\bar{q}, m) \geq U(q, t) = u\}. \quad (\text{F.31})$$

M is a dual function of the utility function, albeit based on time rather than income. Thus, M resembles the cost function in standard economic theory and the proof by Blackorby and Donaldson is easily modified for this case.

Equation (4.20) requires that preferences be homothetic in t . Suppose they are not. Then there exist two health states, (q^0, t^0) and (q^1, t^1) , such that $U(q^0, t^0) = U(q^1, t^1)$. This implies

$$M(q^0, t^0, \bar{q}) = M(q^1, t^1, \bar{q}). \quad (\text{F.32})$$

If preferences are not homothetic in t , then there exists $\lambda > 0$ such that $U(q^0, \lambda t^0) > U(q^1, \lambda t^1)$, implying

$$M(q^0, \lambda t^0, \bar{q}) > M(q^1, \lambda t^1, \bar{q}) \quad (\text{F.33})$$

for all \bar{q} . Note that λ may be chosen to be less than one (the inequality continues to hold).

Now choose $\bar{q} = q^0$ so that

$$M(q^0, t^0, q^0) = t^0. \quad (\text{F.34})$$

From (F.32),

$$M(q^0, t^0, q^0) = M(q^1, t^1, q^0) = t^0$$

and

$$\lambda M(q^1, t^1, q^0) = \lambda t^0. \quad (\text{F.35})$$

From (F.33),

$$M(q^0, \lambda t^0, q^0) > M(q^1, \lambda t^1, q^0). \quad (\text{F.36})$$

By the definition of a minimum, (F.31) yields

$$\lambda t^0 \geq M(q^0, \lambda t^0, q^0). \quad (\text{F.37})$$

Equations (F.35), (F.36), and (F.37) imply

$$\lambda M(q^1, t^1, q^0) > M(q^1, \lambda t^1, q^0). \quad (\text{F.38})$$

By invoking Condition N, the following is obtained

$$M(q^1, 0, q^0) = M(q^1, \lambda 0, q^0) = 0 = \lambda 0. \quad (\text{F.39})$$

Equations (F.38) and (F.39) combine to yield

$$\begin{aligned} M(q^1, \lambda t^1 + (1 - \lambda)0, q^0) &= M(q^1, \lambda t^1, q^0) \\ &< \lambda M(q^1, t^1, q^0) + (1 - \lambda)M(q^1, 0, q^0) = \lambda M(q^1, t^1, q^0) \end{aligned} \quad (\text{F.40})$$

which contradicts the definition of concavity.

Sufficiency:

If $U(q, t) = \phi(a(q)t + b(q))$, then $m(U, \bar{q}) = \bar{a}(q)t + \bar{b}(q)$, which is concave (non-strictly) in t . \square

Lemma 1: Additivity of Preferences

In the two period case (which can be generalized), additivity requires

$$\varphi(q^1 + q^2 | t^1 + t^2)(t^1 + t^2) = \varphi(q^1 | t^1)(t^1) + \varphi(q^2 | t^2)(t^2) \quad (\text{F.41})$$

(where φ denotes the QALY function and the superscripts represent different time periods). This requires

$$U(q, t) = \phi(a(q)t).$$

Necessity:

Recall that

$$U(q, t) = U(\bar{q}, m), \quad (\text{F.42})$$

$$U(q, t) = \bar{U}(m), \quad (\text{F.43})$$

$$\bar{U}^{-1}(U(q, t)) = m \quad (\text{F.44})$$

Since $\varphi(q|t) = \bar{U}^{-1}(U(q, t))/t$, $\varphi(q|t)t' = \bar{U}^{-1}(U(q, t))$ if $t = t'$.

Then the additivity condition becomes

$$\bar{U}^{-1}(U(q^1, q^2, t^1 + t^2)) = \bar{U}^{-1}(U(q^1, t^1)) + \bar{U}^{-1}(U(q^2, t^2)). \quad (\text{F.45})$$

Set $q^1 = q^2$ and incorporate a as a parameter:

$$f(t^1 + t^2) = f(t^1) + f(t^2). \quad (\text{F.46})$$

This is a Cauchy equation with the solution: $f(t) = at + b$. For each q there exists $f(t; q) = a(q)t + b(q)$. Since $f(q, t) = \bar{U}^{-1}(U(q, t)) = \phi(U(q, t))$,

$$U(q, t) = \phi(a(q)t + b(q)). \quad (\text{F.47})$$

However, since $f(q, 0)$ is normalized to be zero, regardless of q , $b(q) = 0$ for all q .

Sufficiency:

If $U(q, t) = \phi(a(q)t)$, then $\varphi(q|t) = \frac{a(q)t}{a(\bar{q})t}$. Summing over two periods:

$$\frac{a(q^1)t^1}{a(\bar{q})t^1}(t^1) + \frac{a(q^2)t^2}{a(\bar{q})t^2}(t^2) = \frac{a(q^1)t^1 + a(q^2)t^2}{a(\bar{q})(t^1 + t^2)}(t^1 + t^2). \quad (\text{F.48})$$

□

Lemma 4: Time Independence:

Necessity:

Time independence requires that

$$\varphi(q^1|t^1)t^1 = \varphi(q^1|t^2)t^1, \quad (\text{F.49})$$

or that $\varphi(q)$ be the same for all values of t (i.e. $\frac{\lambda(m)}{\lambda t} = \frac{m}{t}$ for any positive value of λ). From the definition of φ , this requires

$$U(q, \lambda t) = U(\bar{q}, \lambda m) \quad (\text{F.50})$$

for all $\lambda > 0$. Set $\lambda = 1/t$. Then the problem becomes

$$U(q, 1) = \tilde{U}(q) = U(\bar{q}, m/t) = \bar{U}(m/t). \quad (\text{F.51})$$

Taking the inverse of \bar{U} ,

$$\bar{U}^{-1}(\tilde{U}(q)) = a(q) = (m/t) \quad (\text{F.52})$$

or $m = a(q)t$. But $m = \phi(U(q, t))$, so this implies $U(q, t) = \phi(a(q)t)$.

Sufficiency:

If $U(q, t) = \phi(a(q)t)$, then $m = a(q)t/a(\bar{q})$ and $\varphi(q/t) = m/t = a(q)/a(\bar{q})$, which is clearly independent of t . \square

Appendix G

Empirical Analysis to Chapter 4

G.0.1 Data Description

Data are extracted from four sources: (1) morbidity data are taken from the 1985 General Social Survey (G.S.S., Statistics Canada [1987]), (2) institutional data are taken from the 1986 Health and Activity Limitation Survey (H.A.L.S., Statistics Canada [1990]), (3) QALY values are taken from Torrance et al. (1982) (as reported in Drummond et al. [1987]), and (4) life expectancy data are taken from the 1985-87 life tables (Statistics Canada [1991]).

The G.S.S. covers a stratified sample of all Canadians age 15 and over (i.e. children are excluded), living in the ten provinces (i.e. persons residing in the territories are excluded), and not living in institutions.¹

The health characteristics surveyed in the G.S.S. depend on the age of the respondent, persons over 55 years of age being asked a supplemental set of questions. Thus, two data sets are constructed: one includes all respondents but covers a smaller set of variables, the other includes only older respondents, but has a larger set of variables. Variables chosen include

endurance=1 if a positive response to any of questions 27 (walk), 28 (climb), 29 (carry), or 31 (bend);

¹Families with members over 65 years of age who were either living on Indian Reserves or were members of the Armed Forces are also excluded, although the population weights apparently account for these biases.

=0 otherwise.

endurance2=1 if a severe response to any of the above;

=0 otherwise.

role=1 if a positive response to question 37 (activity limitation);

=0 otherwise.

role2=1 if the response to question 151 is permanently unable to work;

=0 otherwise.

emotion=1 if the response to question 75 is unhappy;

=0 otherwise.

social=1 if, with respect to either contacts or visits, the respondent reports less than the average frequency of encounters in a majority of categories of types of people with whom such contacts can be made (questions 107, 108, 112, 112, 116, 117, 121, 122, 124, 125, 127, 128);

=0 otherwise.

hearing=1 if positive response to question 36 (hearing);

=0 otherwise.

sight=1 if positive response to question 35 (reading);

=0 otherwise.

perception=1 if severe response to either question 35 or 36;

=0 otherwise.

short=1 if positive response to question 17 (sick days);

=0 otherwise.

short2=1 if positive response to question 13 (bed days);

=0 otherwise.

agility=1 if positive response to question 33 (grasp) or 34 (reach);

=0 otherwise.

agility2=1 if severe response to either question 33 or 34;

=0 otherwise.

Additional variables for the over 55 group include

mobility=1 if positive response to question 87 (yardwork) or 91 (light housework);

=0 otherwise.

mobility2=1 if severe response to either question 87 or 91;

=0 otherwise.

selfcare=1 if positive response to question 103 (self care);

=0 otherwise.

selfcare2=1 if severe response to question 103;

=0 otherwise.

Analysis includes the comparison of the above variables with reported satisfaction (question 73). Stratification variables include province, sex, and age. The weight variable is used in the calculation of averages. Observations with inadmissible responses to any of the above (e.g. not stated, did not know, no opinion) are omitted.

These comprise less than .5 per cent of the sample. The data set defined over all age groups has 10739 observations, while the data set defined over the over-55 group has 6688 observations.

G.0.2 Calculation of Satisfaction Estimates

The results of the second chapter suggest a multiplicative functional form should be chosen for estimation.²

Health variables (as described above) are regressed against satisfaction with health. This assumes that all individuals in the sample assign the same meaning to each of the categories (e.g. that very satisfied is very satisfied to everyone). Probit analysis is used because the dependent variable is categorical. This is done with the non-linear procedure in SHAZAM, using the log-density for an ordered probit. Linear estimates are used as starting values, although the final estimates are robust to the choice of initial values. The performance of the estimates based on the whole population are superior to those based on the over 55 data set, even though the latter

²The same conclusion appears to be supported with this estimation, despite slight differences in the variable set considered. This conclusion is drawn from results obtained by imposing additivity on the equation estimated, but deleting observations with multiple health ailments. If these health ailments are additively separable, the resultant parameter estimates should not be significantly different from those based on the entire data set. If additive separability does not hold, then the parameter estimates will differ. (Use of such a procedure is necessary since it is impossible to estimate the equation with higher order terms.) Over all age groups, additive independence is imposed alternately on (1) all variables (no responses are deleted), (2) on the social variables only (all multiples except those involving social ill-health and one other variable are deleted), (3) on the social variables and *within* chronic ill-health (multiples within endurance, agility, and perception are allowed), and (4) on no variables (all multiple responses are omitted). LR tests based on the different parameter estimates indicate additive independence is violated in the first case ($\chi^2 = 216.40$, with 13 degrees of freedom), marginally violated in the third ($\chi^2 = 20.55$), but not violated by the second case ($\chi^2 = 2.154$). These results mimic those found in Chapter 2, and the other structural results found there are assumed to apply here as well (e.g. multiplicative structures cannot be refuted). Because more observations entail more precise estimates, the second case is adopted for estimation (8407 observations are left). The performance of the parameter estimates from the third case are inferior at the stage when they must be linked to QALY values.

contains two extra explanatory variables.³ Thus, the all-inclusive data set is adopted.

The data set used for estimation is then the one for all age groups with observations having more than one ill-health state (social ill health excluded) deleted. This assumes additive independence only of social ill health from the other ill health categories. The estimate associated with any other ill health variable is then a measure of the marginal disutility of moving to that health state from a state of perfect health and is free of any interaction effects from any co-morbid state (e.g. the parameter value for endurance measures the change in satisfaction, or rather the probability associated with reporting this level of satisfaction, when an individual moves from a state of perfect health to one characterized by some endurance impairment; it does not indicate the change in satisfaction when someone who also has a perception impairment acquires an endurance impairment – these values are the same only if preferences are additively separable over endurance and perception). The estimates based on the general data set with additive independence imposed on social ill-health are reported in Table G.1. All parameter estimates have the expected sign (since ill-health should contribute to health dissatisfaction). Notice that, given how the severe variables are defined, the marginal probabilities associated with severe ailments are found by adding the coefficients for the ailment being present and severe.

The estimates are checked for variation across identifiable groups to verify the constancy of preferences. This is done by regressing calculated residuals against the independent variables after sorting observations by group characteristics. The set of stratification variables used is more limited than in Chapter 2, since the focus here is to establish whether preferences vary only across the groups whose health is being

³In the latter case, the procedure sometimes fails to converge, the estimates are more sensitive to the inclusion of certain observations, and the estimates are occasionally of the “wrong” sign. This may be due to misspecification of the extra variables, multicollinearity with other variables, or age dependence in the estimates. Correlation statistics and the performance of the parameter estimates lends support to the latter two hypotheses.

Table G.1: Satisfaction Function Estimates (t-statistics in brackets)

variable	all	women	men
constant	-.107 (-6.54)	-.093 (-4.13)	-.120 (-5.10)
endurance	.402 (8.80)	.337 (5.97)	.539 (6.83)
endurance2	.077 (.440)	.314 (1.52)	-.499 (-1.47)
role	.685 (8.57)	.627 (5.51)	.733 (6.47)
role2	.260 (1.40)	.155 (.648)	.476 (1.54)
emotional	.749 (7.79)	.650 (5.02)	.873 (6.00)
social	.061 (1.54)	.034 (.589)	.083 (1.54)
hearing	.144 (2.31)	.336 (3.33)	.032 (.404)
sight	.191 (1.88)	.132 (.974)	.272 (1.78)
perception	.078 (.283)	-.027 (-.068)	.134 (.344)
short	.358 (6.10)	.425 (5.73)	.241 (2.49)
short2	.259 (4.07)	.251 (3.11)	.254 (2.40)
agility	.104 (1.05)	-.022 (-.168)	.279 (1.82)
agility2	1.08 (2.58)	1.41 (2.79)	.293 (.363)
a_1	1.61 (68.9)	1.64 (50.7)	1.58 (46.4)
a_2	2.65 (48.1)	2.60 (36.9)	2.75 (30.1)
LL	-7368.544	-3923.268	-3431.698

compared where such variability would affect the conclusions drawn. The stratification variables used are sex, age, and province. Tests indicate there is some variation in each group. For age, significant variation occurs in the constant (as is expected since satisfaction with health rather than satisfaction with morbidity is reported and the former probably contains a longevity component that is related to age), but the relationships over the morbidity variables do not, suggesting preferences over morbidity are constant across age groups. Preference variation across provinces cannot be assessed because of insufficient sample size in the smaller provinces. Regional subaggregates reveal little variation. Preferences do vary by sex. This is confirmed by the re-estimation of the satisfaction equations by sex (see Table G.1). LR tests indicate the differences are statistically significant ($\chi^2 = 68.04$ between the men and women with 16 degrees of freedom).

G.0.3 Calculation of Expected Satisfaction

The estimates from the probit equation for the general data set with additive independence imposed on the social ill-health variables are used to generate expected dissatisfaction ($D(q)$) for each single dimensioned health state (i.e. when only one characteristic indicates poor health). These are converted to utility values by the following negative affine transformation

$$S(q) = \frac{D(q) - D(all\ q)}{D(no\ q) - D(all\ q)} = \frac{D(q) - 3.947410}{-2.44372}, \quad (G.1)$$

where $S(q)$ is the estimated satisfaction associated with health characteristic q , $D(q)$ is the estimated dissatisfaction, $D(all\ q)$ is the dissatisfaction associated with the poorest possible health state, and $D(no\ q)$ is the dissatisfaction associated with the best possible health state. For women, the transformation is

$$S^F(q) = \frac{D^F(q) - 3.982166}{1.508255 - 3.982166}, \quad (G.2)$$

Table G.2: Marginal Disutilities (taken at perfect health)

variable	all	women	men
endurance	.089	.072	.130
endurance2	.107	.147	.009
role	.158	.141	.182
role2	.226	.180	.317
emotional	.174	.147	.221
social	.013	.007	.018
hearing	.015	.072	.007
sight	.040	.027	.063
perception	.075	.097	.104
short	.078	.093	.055
short2	.141	.153	.118
agility	.022	.001	.064
agility2	.292	.377	.139

whereas for men it is

$$S^M(q) = \frac{D^M(q) - 3.777051}{1.498834 - 3.777051}. \quad (\text{G.3})$$

These satisfaction values are reported in Table G.2.

To find the values associated with the multidimensional health states, the parameter restriction associated with multiplicative utility structures is invoked

$$(1 + \lambda) = \Pi(1 + \lambda\lambda_i) \quad (\text{G.4})$$

where λ_i is the marginal probability associated with the worst level of characteristic i (i.e. endurance2, role2, social with emotional, perception (with both blindness and deafness), short2, and agility2).⁴ This is consistent with the procedure adopted by Boyle and Torrance (1982). The restriction is solved for λ . Although a fifth

⁴Note that this procedure may be somewhat biased since estimated rather than actual values are used. The unbiased version (in the two characteristic case) requires

$$1 + \lambda = 1 + \lambda(\lambda_1^e + \lambda_2^e) + \lambda^2(\lambda_1\lambda_2)^e$$

(where “e” denotes the expectation operator), whereas the estimated version gives

$$1 + \lambda = 1 + \lambda(\lambda_1^e + \lambda_2^e) + \lambda^2\lambda_1^e\lambda_2^e.$$

order polynomial, only one root satisfying rationality restrictions was found for each equation (all data, men, and women). The only non-zero real root for the whole data set is -.8312363. The only non-zero real root for the male equation is -.2345702, whereas for the female equation there are three non-zero real roots, but only one falls in the bounds dictated by multi-attribute theory: .2010460.

These values are then used to calculate a second-order approximation of a multiplicative utility function:

$$U(q_1, \dots, q_N) = \sum_{i=1}^N S(q_i) + \sum_{i=1}^N \sum_{j>i}^N \lambda S(q_i) S(q_j). \quad (G.5)$$

Notice that the same root value of λ is used regardless of the severity level of the characteristic (as is done by Torrance et al. [1982]) and that only mutually exclusive characteristics are combined in this way.

G.0.2 Linkage to QALY Values

Because the expected utility values calculated above are not linked to the value of life, against which they will eventually be compared, it is necessary to transform them so that they map into the QALY time trade-off interval. Since the time trade-off function is assumed to be a monotonic function of utility over morbidity conditioned on time, it is necessary to either match QALY and satisfaction time frames or assume

Since

$$(\lambda_1 \lambda_2)^e = \lambda_1^e \lambda_2^e + COV(\lambda_1, \lambda_2),$$

the estimated version yields

$$\lambda(\lambda_1^e + \lambda_2^e - 1) = \lambda^2(\lambda_1 \lambda_2)^e - COV.$$

Let $a = \lambda(\lambda_1^e + \lambda_2^e - 1)$ and $b = \lambda^2((\lambda_1 \lambda_2)^e - COV)$. Solve for λ :

$$\lambda^* = (-a \pm \sqrt{a^2 + 4b})/2b = 0, a/b.$$

Since $\lambda^* = 0$ is the additive case, assume $\lambda^* = a/b$. Differentiate with respect to COV to get $\frac{a}{b^2} \frac{\partial b}{\partial COV}$. Since $b^2 > 0$, $\frac{\partial b}{\partial COV} \leq 0$, $\frac{\partial \lambda}{\partial COV} > / < 0$ as $a < / > 0$ (i.e. if the agent is risk averse, the bias is negative; if the agent is risk seeking, the bias is positive.)

QALYs are independent of time. Since the former assumption cannot be assessed with the G.S.S. survey (no time dimension is given on the satisfaction response), it is necessary to assume time independence to proceed.

The monotonic function relating expected utility with QALYs is recovered by matching morbid states surveyed in the G.S.S. with those for which QALY values (obtained by time trade-off) exist. These values are taken from Torrance et al. (1982) (found in Drummond et al. [1987]) and include

endurance=P2

endurance2=P3

role=R2+R4/2

role=R3+R5/2⁵

social=S2

emotional=S3

sight=H6

hearing=H3

perception=H8

Perfect health is matched to the absence of any reported health impairments for the tenth match.

⁵For both P and R, the G.S.S. provides only partial information (i.e. a subset of the characteristics actually used in the Torrance et al. valuation exercise). Since the correlation of the missing and observed characteristics is unknown, the values of P and R are chosen to most closely approximate the ordering of health states by expected utility. For state P, this requires assuming endurance and mobility are uncorrelated, whereas for R it requires that selfcare be inhibited by half the respondents who are role dysfunctional (only the one mid-point is considered).

The QALY values for the states above are derived from middle-aged parents living in and around Hamilton, Ontario. The linkage generates biased values if the preferences of these parents differ from those of the general Canadian population. To ascertain if this is the case, preference variation between middle-aged Ontario residents and the Canadian population is tested for using a Hausman-Wise (1978) residual test (the calculated errors for this group are regressed on the explanatory variables). No significant differences are found (no t-test on any of the explanatory variables is significant, and the overall F-test is only .629 (with 13 degrees of freedom), far below the critical level), so the linkage is performed on the parameter estimates for the whole G.S.S. sample.

The QALY and expected utility values associated with these states are then used to uncover the relationship between the two value scales. The linkage is difficult to make since both the functional form and the parameters must be estimated and there are only ten common observations, all skewed towards perfect health. Four functional forms are evaluated: a basic linear, a translog (first and second order), a quadratic (first and second order) and a Box-Cox. In both cases where they are tried, the performance of second order approximations is found to be inferior to first order approximations (t-tests indicated none of the second order terms is statistically significant and the second order approximations are far more sensitive to the inclusion or exclusion of matched observations). The non-linear forms, the Box-Cox and the double log (first order translog), dominate the linear forms (the linear and first order quadratic). Of the two non-linear forms, the double log is chosen because (1) the Box-Cox estimates are less stable, especially when observations are omitted, (2) model selection statistics are generally better for the double log, and (3) the double log provides the intermediate values of all the functional forms tried (a plot of the Box-Cox, double log, and linear functions are virtually identical in the range where data

observations exist, and differ only in the extremely bad states; the Box-Cox gives the highest values to these states, the linear the lowest).⁶

The logarithmic transformation used is $\hat{\varphi}(q) = \dot{U}(q)^{1.4321}$ (standard error on the parameter is .20840).

Because the preferences of men and women for different health states apparently differ, this procedure has to be repeated for men and women separately. Unfortunately, the mean QALY values reported are not broken down by gender. This creates the potential for two sorts of bias: first, a linear average of the estimated QALY may misrepresent the true average QALY and, second, the weights used in the two value indexes may differ.

The Case Against Common Transforms

Consider the case where mean QALYs are regressed against average utility, but the proportion of men and women in the two groups is the same, expressed as a^M and a^F respectively. If the true relationship is described by f , then the regression involves

$$a^m \varphi^M + a^F \varphi^F = f(a^M U^M + a^F U^F) \quad (\text{G.6})$$

$$a^m f^M(U^M) + a^F f^F(U^F) = f(a^M U^M + a^F U^F). \quad (\text{G.7})$$

But this is a Cauchy equation with the solution

$$f^i(U^i) = aU^i + b. \quad (\text{G.8})$$

Suppose this is not the case, but that f is in fact strictly convex (as the estimated coefficients suggest). Then

$$\hat{\varphi}(a^M U^M + a^F U^F) > a^M \varphi^M + a^F \varphi^M \quad (\text{G.9})$$

⁶This decision is supported by Torrance (1976b) who found time trade-off values relate to category scale values (which resemble the G.S.S. responses in design) by a logarithmic function.

(i.e. the estimated QALY overstates the true average QALY, for any proportion, regardless of which gender has a higher utility for any particular state). The greater the difference in men's and women's preferences, the greater is this bias.

The Case Against Specific Transforms

Conversely, consider the case where the proportion of men versus women differs between the two groups. Even if QALYs and utilities are cardinally related, bias is present whenever preferences differ (the estimated QALY is biased upwards if the utilities are based on fewer men than the true QALY and men value the state more than women, et cetera). Since Torrance et al. (1982) do not provide a breakdown by sex of the respondents, it is not possible to assess the impact of this situation. For the purposes here, it is assumed that their random sample of parents reasonably approximates the proportions found in the G.S.S. after weighting. This does pose a problem when the specific transforms are calculated since the proportional representation on the utility side is (0, 1), while it is closer to 50-50 on the QALY side. In this case, the bias depends entirely on the differences in utility values between the sexes.

Since it is not clear that either of the above two methods generates less biased transforms, both are estimated. It is obvious that both methods generate some bias (although the specific transform case probably generates less than the common transform case).

As for the grouped data, the logarithmic form is found to perform the best for both the male and female equations. Parameter estimates do differ:

$$\hat{\varphi}(q)^{JT} = \hat{U}^{JT}(q)^{1.4321},$$

$$\hat{\varphi}(q)^M = \hat{U}^M(q)^{1.1568},$$

$$\hat{\varphi}(q)^F = \hat{U}^F(q)^{1.7578},$$

(with standard errors on the parameters of .20840, .19595, and .22186 respectively).

G.0.5 Comment on Institutional Data

The most serious omission from the G.S.S. as a source of health status data is its exclusion of institutionalized persons. Data on the disabled population residing in institutions are available for Canada and the provinces by age and sex in the H.A.L.S. (1990). Three adjustments are necessary before these data can be incorporated into the above morbidity data. First, the groups excluded from the G.S.S. (those under 15 years and those residing in the two territories) have to be deleted from the H.A.L.S. figures. Second, because the H.A.L.S. is based on a survey taken the year after the G.S.S., the population figures have to be scaled down to 1985 levels. This is done by taking the population base of the G.S.S. (total number of observations in the sample multiplied by the average weight: 11200×1756) and dividing by the total population in the H.A.L.S. (after the above deductions have been made). This figure is used to scale down the H.A.L.S. data (this assumes a constant proportional relationship between the 1985 and 1986 populations under 15 years, residing in the territories and living in institutions). These adjusted populations give the appropriate weights to be used when the institutional data are combined with the G.S.S. morbidity data.

The third adjustment is to link the morbid state (being institutionalized) to the QALY value scale. Since these states are external to the G.S.S. preference function, the procedure used for the G.S.S. states cannot be followed. Instead, values for hospital confinement found in Sackett and Torrance (1978) are used. The value for the state depends critically on the duration of the state, which is not available in the H.A.L.S. (it being a point in time survey like the G.S.S.). It is assumed the average duration must lie between 3 months (a QALY value of .56) and 8 years (a QALY value of .33). These values, combined with the population weights above, are then incorporated into the QALY averages.

G.0.6 Calculation of QALY Averages

QALY averages are calculated from the estimated QALY values, using the G.S.S. weights (and the institutional weights when the H.A.L.S. is incorporated). In the G.S.S. data set, this is done by five year age groupings, but when combined with the H.A.L.S. data set, only ten year demarcations are available. Because morbidity appears to increase with age, but the number of people alive does not, less severe morbid states receive a higher weight in these averages than the more severe states. This bias worsens the more years on which the average is based. Thus, the 5 year groupings should give a lower average morbidity level than the 10 year groupings, although both are biased upwards.

Because of data limitations (longitudinal data are unavailable), arithmetic averages are used above. These mask the degree of inequality in the population's morbidity and health statuses. Although morbidity averages are misleading, because they do not incorporate length of life, it is still interesting to observe how inequality aversion affects the results. For the adult, non-institutionalized population, over all ages, the arithmetic mean of estimated QALYs (based on common preferences) is .917 (this reflects the position of inequality neutrality). The geometric mean (reflecting Cobb-Douglas social preferences and some inequality aversion) is .903 and the minimum value (reflecting Rawlsian preferences) is .203. If everyone lives a seventy year lifespan, the inequality aversion in the Cobb-Douglas increases the willingness to give up longevity in order to eradicate morbidity by 1 year, nearly a twenty per cent increase over the inequality neutral case. As inequality aversion increases, so does this difference. This underscores the need for adequate longitudinal data, since without them, the health index is forced to adopt inequality neutral ethics, despite the obvious inequalities present in society. This is particularly relevant in health policy which can

often effect discrete changes in society's health status profile.

G.0.7 Life Expectancy

Life expectancy data are taken from Statistics Canada (1991), since the population base for these figures most closely approximates that of the G.S.S. (1985-1987 versus 1985). Three adjustments are made so that the bases are as comparable as possible.

First, life expectancy is conditioned on living to the fifteenth birthday, rather than being born, i.e.

$$\begin{aligned} E(t|t \geq 15) &= .5P^D(15|15) + P^S(15|15) + \\ &P^S(15|15)(.5P^D(16|16) + P^S(16|16)) + \dots + \\ &P^S(15|15)\dots P^S(T-1|T-1).5P^D(T|T), \end{aligned} \quad (\text{G.10})$$

where $P^D(t|t) = \text{probability of dying at age } t \text{ given survival to the } t\text{th birthday}$ ($= \frac{D_t}{L_t}$ where D_t is the number of people who die in the t th year and L_t is the number of people who are alive at the beginning of the t th year), $P^S(t|t) = \text{probability of surviving the } t\text{th year given survival to the } t\text{th birthday}$ ($= \frac{L_t - D_t}{L_t}$), and T is the maximum age.

By cumulative laws of probability

$$E(t|t \geq 15) = .5P^D(15|15) + P^S(15|15) + .5P^D(16|15) + \dots + .5P^D(T|15), \quad (\text{G.11})$$

where $P^D(t|15) = \frac{D_t}{L_{15}}$ and $P^S(t|15) = \frac{L_{15} - \sum_{i=15}^{t-1} D_i}{L_{15}}$. Since T is not given in the tables, life expectancy for the last period (80+) has to be calculated

$$\frac{L_{80}}{L_{15}} E(80|80) = E(80|15) \quad (\text{G.12})$$

(i.e. life expectancy at age 80 is scaled so that the probabilities are conditioned on living to the 15th birthday and not the 80th).

If life expectancy from birth is used, the health status index is biased upwards since no morbidity is reported for the under 15 group. The measure proposed is necessarily incomplete, but unbiased.

It is assumed that people die at the mid-point of the t th year on average (hence, P^D is multiplied by .5). Since people over fifteen generally die at an increasing rate, this involves some positive bias which can be minimized by choosing the time interval to be as small as possible (i.e. one year) and aggregating these into the larger intervals by simple summation. This is the second adjustment. Thus, the number of years of expected life between any two ages is calculated as the sum of $P^S(t|15) + .5P^D(t|15)$ over all t lying within the interval. Intervals are 5 years when G.S.S. data alone are used, and 10 years when H.A.L.S. data are used as well.

Third, the life expectancy figures for Canada incorporate data from the two territories. Since the effect of these persons on the national figures cannot be isolated, the analysis is performed for each province as well as for the country. The provincial figures are uncontaminated by the territories' influence, but are subject to the vagaries of small sample sizes in some cases. Since the effect of the territories' different mortality profiles is apt to be dominated by the provinces' larger populations, the bias should be quite small.

G.0.8 QALT Calculations

Life expectancy at each age is adjusted by the mean QALY value for persons of that age:

$$HSE = \sum_{t=15}^{80} [P^S(t|15) + .5P^D(t|15)] \left[\sum_{i=1}^{N_t} \hat{\varphi}(q_{i|t}) / N_t \right], \quad (G.13)$$

where N_t is the number of people in age category t .

Standard errors are estimated using Rothenberg's (1984) Taylor's series expansion

over the non-linear function of random variables used to calculate the quality adjusted life expectancy ($\sum_t Pr(t|15)b_i^\alpha q_{i|t}$, where b_i s are the satisfaction values given in Table G.2, the α s are the power transforms used to convert the satisfaction values to a time trade-off consistent scale, and the $q_{i|t}$ s are the morbid states endured by the people surveyed in age group t – note that the probability of survival is treated as parametric since it is based on the entire population of interest). Such a method approximates the true standard errors asymptotically if the estimators of the random variables used in the function are consistent and asymptotically normal. It is assumed in these calculations that the estimators are independent of one another.

A problem of bias may result from the fact that the life tables are based on a steady state estimate of the population, while the G.S.S. is based on the actual population. Because of the post-war baby boom, this means the G.S.S. has more observations in the younger and middle-aged segments of the population. If the proportion exceeds the probabilities at the younger years when people are generally healthier, the estimated health status is biased upwards. However, since the average is taken over 5 year intervals, there does not seem much opportunity for such a bias to have an effect.