

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

**ECONOMIC EVALUATION
OF HEALTH CARE:
CAUTIONS FOR THE
DEVELOPING COUNTRY
CONTEXT**

By Jolene Skordis

A thesis submitted in partial fulfilment of
the requirements for the degree of

Masters in Economics and Demography

University of Cape Town

2002

Approved by _____

Program Authorized
to Offer Degree _____

Date _____

ACKNOWLEDGMENTS

For their financial support during my Masters degree, I would like to thank the Mellon Foundation and the UCT Scholarships Committee.

For their professional support and advice, I would like to thank Nicoli Natrass; Edina Sinanovic; Di McIntyre; Debbie Bradshaw and Murray Leibbrandt.

For his personal support, I would like to thank my lovely husband, Lyndon, for putting up with the late nights and teaching me how to be patient with PDF files.

University of Cape Town

Abstract

ECONOMIC EVALUATION OF
HEALTH CARE: CAUTIONS FOR THE
DEVELOPING COUNTRY CONTEXT

by Jolene Skordis

Supervisor: Professor Nicoli Nattrass (Department of Economics)

Health economics has expanded enormously as a sub-discipline in the last four decades, drawing primarily on the theoretical foundations of welfare economics. The toolkit for the economic evaluation of health care now extends from the humble cost-minimisation exercise, through cost effectiveness measures, to the more complex cost utility or cost benefit models. These methodologies have differing strengths and drawbacks. This paper evaluates those attributes on both the practical and theoretical dimensions.

On the practical dimension: The developing country context differs from the wealthier country context in a number of ways. This paper considers the differences in resource constraints and the differences in health priorities and asks to what extent the methodology is able to accommodate these variations.

On the theoretical dimension: Few health care evaluations are conducted in a Pareto Optimal world. This paper considers the extent to which the welfare economic foundations of a methodology are successfully imported into its construction, and then how that foundation translates into its practical application.

TABLE OF CONTENTS

Chapter One: Introduction	2
An important role for health economics.....	2
Selecting a Methodology	7
Chapter Two: The Theoretical Foundations of Health Costing	9
The social welfare Foundations of Health Costing.....	9
How can the Theory of the Second Best Aid Economic Evaluations in Developing Countries?	11
Making the Best of a Second Best Situation In Theory	12
The Positive Versus Normative Trade-off In Practice.....	12
Some final thoughts on the theory of the second best.....	14
Chapter Three: Cost Minimisation – A Rarely Used Tool	15
Cost Minimisation and Output Maximisation.....	15
Cost Minimisation Applied in the Developing Country Context	15
Shortcomings of the Cost Minimisation Methodology.....	20
A Final Word on Cost Minimisation.....	21
Chapter Four: Cost Effectiveness Analysis - An Inadequate Reflection of Social Welfare Ideology.....	23
Introduction.....	23
The apparent Social welfare foundation.....	24
The Individual Case: A Sound “First Principles” Foundation.....	25
The Societal Case: A “First Principles” Imposter.....	28
An Additional Criticism of the CEA Methodology.....	30
A Final Word On Cost Effectiveness Analysis	31
Chapter Five: Cost Utility Analysis – Is There a Way to Reflect Accurately the Utility of Health Care?	35
Why do we need a composite indicator of burden of disease?	35
What is a DALY?	38
Examples of the leading causes Of Mortality and Morbidity Underlying DALYs around the world.....	39
Constructing a DALY	41
DALYs at work – the case of HIV in South Africa	51
HIV Treatment in the Current Political Climate	51
Back to Basics: A Review of ARVs in the HAART Context	52
Devising a unit cost of HAART treatment: An Overview of the Assumptions	54
A final word on DALYs.....	63
Chapter Six: Cost Benefit Analysis – The Human Capital Approach versus Willingness to Pay	66
Introduction.....	66
The Human Capital Approach.....	66
Willingness to Pay	68
Alternative Approaches to Measuring Benefit	72
The Net-Cost Methodology	73
Conclusion	76
Chapter Seven: Conclusion	78

AN IMPORTANT ROLE FOR HEALTH ECONOMICS

Health economics has expanded enormously as a sub-discipline in the last four decades, drawing on the theoretical foundations of welfare economics and the more applied experience of environmental economics and public finance analysis. The toolkit for the economic evaluation of health care now extends from the humble cost-minimisation exercise, through cost effectiveness measures, to the more complex cost utility and cost benefit models. Techniques have evolved to supplement clinical data with survey data from the patient and cost data from the provider. Composite measures have been devised in an attempt to capture the burden of disease and other non-monetary gains from health expenditure. We have developed a better understanding of the health care provider as an economic and social entity.

Unfortunately, very little attention has thus far been paid to the accuracy and suitability of these measurement techniques in various contexts. What little has been written applies most particularly to the case of wealthier countries, despite the argument that lower-income countries have the greatest need to optimise their allocation of scarce resources. This paper considers the first principles foundations of economic evaluation and the particular conditions that may help or hinder the economic evaluation of health care in a developing country context.

Health expenditure comprises a significant and growing proportion of gross domestic product in all countries. In the last four decades the percentage of GDP expended on health has more than doubled in most higher income countries (see *Table One*).^{74,91} In sub-Saharan Africa however, the picture is slightly different. The 650 million people in sub-Saharan Africa have lower life expectancies and higher age-adjusted mortality rates than the rest of the world⁹⁷,

implying more serious health challenges. In 2000/2001, combined public and private health spending in South Africa was estimated at 4.2% of GDP*, down from the 5.6% estimated by Peters *et al*⁶⁸ for 1990. On average in the sub-Saharan region, total government expenditure comprised 25.5% of GDP between 1990-1996. Of this, Education comprised 3.5%, Defence comprised 3% and Health comprised only 1.7%.⁶⁸ The provision of health services consumes substantial resources and the countries with the most dire health problems allocate less of their available resources to health care.

Table One

TOTAL HEALTH EXPENDITURES AS A PERCENTAGE OF GDP				
	1960	1970	1980	1990
EU	3.9	5.2	7.1	7.6
Australia	4.5	5.7	7.4	7.5
Canada	5.5	7.1	7.4	9.0
Japan	2.9	4.4	6.4	6.5
UK	3.9	4.5	5.6	6.1
USA	5.3	7.5	9.3	12.4

Source:^{74,91}

The World Health Organisation's Commission on Macroeconomics and Health has recently published a report stating that the link between the overall health of a population and a country's economic growth is far stronger than previously thought.^{91,97}

* Calculated using the rand-dollar exchange rate of 0.087642 on the 30th of January 2001

The same WHO report identifies the main causes of avoidable deaths in low-income countries as: HIV/AIDS; malaria; tuberculosis (TB); childhood infectious diseases; maternal and peri-natal conditions; micronutrient deficiencies; and tobacco-related illness. The key phrase is “avoidable deaths” and herein lies the problem: In what practical sense are these avoidable when their eradication (or reduction) requires a substantial economic investment over and above what is already being contributed by the affected countries? Additional foreign aid could perhaps go some way towards alleviating these resource constraints but it is unlikely that aid alone would be able to fund all of the necessary programmes in all of the needful countries. Decisions will need to be made: which countries warrant the investment and which programmes within those countries?

The question of “which programmes within those countries” raises a particular dimension of project evaluation. Between the developed and less developed world, resource constraints may differ but so too do health care priorities. For example: HIV is an impending health care catastrophe in South Africa (where prevalence among adults 15-49 stood at around 20%^{90,91} at the end of 2001) in Botswana (where prevalence among adults 15-49 stood at 38.8%^{90,91} at the end of 2001) and in a host of other countries in sub-Saharan Africa. However, in countries such as the USA, UK and Germany, prevalence levels among adults 15-49 were all below 1%^{90,91} at the end of 2001 (see *Figure One*).

Figure One: Global HIV Prevalence⁹⁰

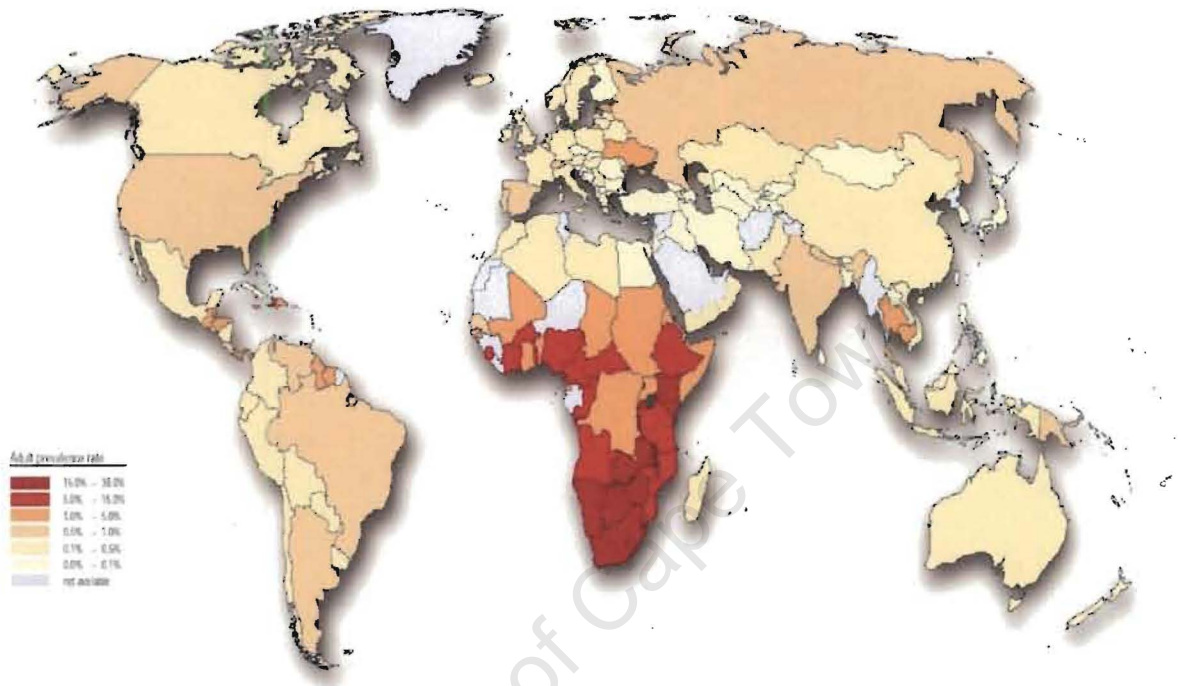
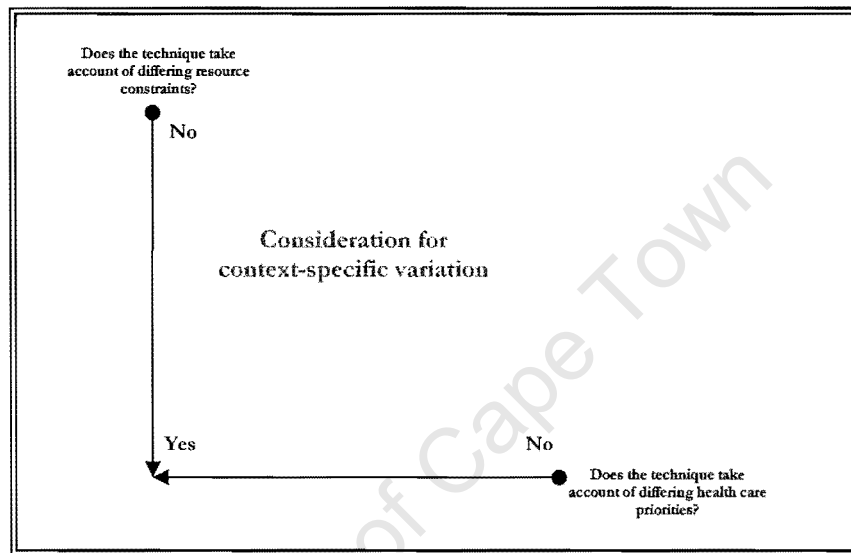


Figure Two over the page illustrates two key areas of difference between health care planning and delivery in the developed versus less developed contexts. The first is the difference in resource constraints and the second is the difference in health care priorities. For example, not only is the GDP of South Africa less than the GDP of the United States⁸⁷, but health spending comprises a smaller percentage of GDP in South Africa than it does in the US⁶⁸. Similarly, the prevalence of HIV in the United States is very different to the prevalence of HIV in South Africa⁹⁰ - suggesting different health care priorities between the two countries. Any methodology intended to measure the effectiveness of health care planning and delivery in a less-developed country needs to be able to take account of these differences. The question that will be asked of each methodology evaluated in this paper will be: “Does this methodology allow for

the consideration of context-specific variation in resource constraints and health care priorities, either in the inputs or the findings?”

Figure Two: Assessing the methodology’s capacity to consider context-specific variation in resource constraints and health priorities.



The two dimensions of the grid in *Figure Two* are deliberately constructed as interdependent. This interdependence was illustrated above by the WHO’s list of illnesses causing “avoidable deaths”. The deaths are avoidable in the absence (or relaxation) of resource constraints, and the list consists largely of illnesses one would not expect to find in wealthier countries (TB is a particular example here). Hence resource constraints can drive the variation in health priorities between “rich” and “poor” countries. However, they are certainly not the only driver, for instance the prevalence of malaria (and malaria-related deaths) could be considered to be environmental as well as resource driven. The qualitative framework in *Figure Two* will be used throughout this paper to assess the suitability of common health evaluation techniques for developing country contexts.

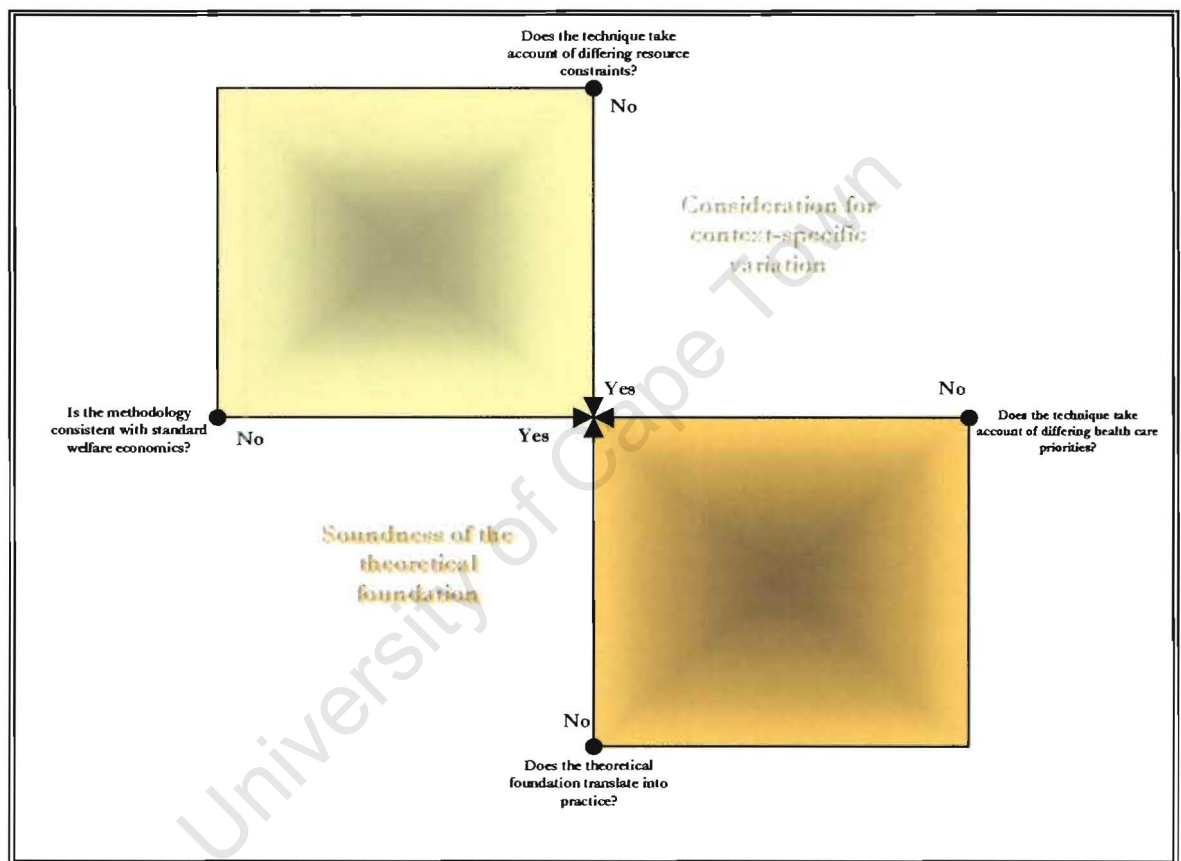
SELECTING A METHODOLOGY

One of the first steps in conducting an economic evaluation is to understand the nature of the question or problem at hand. Cost effectiveness analysis (CEA), cost benefit analysis (CBA), cost minimisation (CM) and cost utility analysis (CUA) all require some valuation of cost and some valuation of benefit. The primary point of divergence between the methods is in the way that benefit is measured. Cost minimisation holds the benefit constant between two interventions and simply attempts to minimise costs.^{48,82} In cost effectiveness analysis, benefit is measured in terms of a specific health outcome - for example the number of malaria cases prevented or the number of diarrhoeal infections cured.^{34,60,70,81} In cost utility analysis, the measurement of benefit is in terms of utility as captured by a disability-adjusted life year (DALY), a quality adjusted life year (QALY) or a healthy life year equivalent (HYE).^{20,82} Cost benefit analysis, often considered the gold standard of economic evaluation, measures the net monetised benefit to a society (or group within a society) of an intervention.^{15,91} Cost benefit analysis is the only methodology that appears to be entirely consistent with textbook welfare economics.^{91,15,20,40,88}

This consistency with economic welfare theory presents a further dimension for the assessment of economic evaluation of health care. To what degree are these methods based on first-principles economic theory? And to what degree does that theoretical foundation translate into practice? It is necessary to distinguish between the theory and the practice because the move from theory to practice could be more difficult in the developing country context and certainly fraught with different challenges. The limited availability of data on health care utilisation (from the patient's perspective) and the cost of health care delivery (from the provider's perspective) are just two examples of how the practical application of economic theory might be complicated in the developing country context. *Figure Three* over the page adds this theoretic dimension of the assessment to the

context-specific dimension illustrated in *Figure Two*. Again, the dimensions of the theoretic “matrix” are deliberately constructed as interdependent criteria. The closer a methodology moves to the centre of the diagram, the better the methodology fares on all criteria.

Figure Three: A matrix for the assessment of the economic evaluation of health care



Before initiating the critical assessment proposed, it is necessary to establish what the terms “welfare economics” or “first-principles economic foundations” mean. The assessment matrix outlined in *Figure Three*, will then serve as a framework to evaluate the soundness of economic measurement techniques in the developing country context.

*Chapter Two: The Theoretical Foundations of
Health Costing*

THE SOCIAL WELFARE FOUNDATIONS OF HEALTH COSTING

At the core of welfare economics is the familiar problem of utility maximisation under constraint.^{77;78;86;91} It is generally assumed that economic actors behave rationally and that preferences are revealed through the actions of the individual.^{17;77;91;94} This enables the State, as the likely coordinator of public welfare, to maximise the utility of society as a whole through the optimisation of an appropriate social welfare function.^{86;91;93;94} This is theoretically achievable through the use of the Pareto rule for welfare improvements i.e. if you can make one member of society better off without making anyone else in society worse off, then you have achieved a Pareto welfare improvement.^{17;43;91;94} If no such improvements are possible then the society has achieved a state of Pareto optimality.

A health economist is likely to find fault with the application of this premise for a number of reasons. Firstly, is this the role that the State should be playing?^{78;91} Secondly, how do we really know the implications of a health intervention at the margin?^{1;20;78;91} Thirdly, the theory supposes that individual's preferences are revealed through the exercise of choice.^{17;91;94} In many markets, preferences cannot be revealed because of institutional failure, market failure or a lack of individual capacity. Can an unborn child reveal a preference for his HIV positive mother to receive anti-retroviral treatment?

Finally, the existence of a single global maximum - i.e. a single optimal solution to a maximisation problem - requires strict convexity of the utility function. In many other markets this does not pose a problem, as the law of diminishing marginal returns to consumption will produce a convex utility curve.^{91;94}

However, in the case of many medical interventions, the law of diminishing marginal returns does not always hold. Consider a patient on chronic asthma medication. As long as the patient continues to take the medication he or she will enjoy good health. The body does not build up a resistance to the medicine over time nor does it display any adverse side effects. Failure to take the medication will result in disutility – dizziness, shortness of breath and possibly even death. This case describes constant rather than diminishing marginal utility. The solution to an optimisation problem such as this will not yield a single global maximum. From a public health perspective, another source of non-convexity is the existence of externalities such as those which arise when an epidemic is controlled through inoculation.

In their seminal paper, Lipsey and Lancaster^{46,91} address the problem of reaching an optimal solution in a world where the theoretical assumptions of standard welfare economics do not apply. They coined the phrase “the second best world” when referring to the world outside of the optimal solution. This paper, and indeed this problem in general, has sparked a wealth of literature on the subject, some of which has important implications for the economic evaluation of health care interventions, particularly in the developing country context.

HOW CAN THE THEORY OF THE SECOND BEST AID ECONOMIC EVALUATIONS IN DEVELOPING COUNTRIES?

According to Lipsey and Lancaster (1956),^{46,91} the “General Theorem of the Second Best” states that if a constraint is introduced into a general equilibrium system that prevents the attainment of one of the Paretian conditions[†], then the other Paretian conditions (which may still be attainable), are no longer necessarily desirable.

Real markets are seldom likely to satisfy all of the conditions for Pareto optimality. If Pareto optimality is both a necessary and sufficient condition for welfare maximization, what is a sufficient condition for at least a welfare improvement when a market or society is not at the maximum?^{46,91} What solutions can economic theory provide?

There appear to be two primary schools of thought regarding Lipsey and Lancaster’s article. One interpretation is concerned with the idea that outside of Pareto Optimal equilibrium, we can say nothing positively about optimal resource allocation.^{16,91} The second school of thought insists that economic theory can still be of some assistance to the decision-maker by drawing on the measurement of deadweight loss.^{16,86,91} Each of these views obviously has quite different implications for policy design.

[†] These conditions require: perfect competition in all sectors of the economy so that marginal cost will necessarily equal price; there should be no external economies or diseconomies of scale; there should be no divergence between private and social cost anywhere in the economy. ^{47,91}

MAKING THE BEST OF A SECOND BEST SITUATION IN THEORY

Davis and Whinston¹⁶ pursued the idea of using normative decision rules by defining a Pareto optimum problem as one in which all normative behavioural rules can be determined so that the solutions of the system achieve a vector maximum i.e. a set of possible maxima rather than a single global maximum. They define a second best problem as one in which at least one of the normative behavioural rules is non-trivially specified, cannot be changed, and in which any deviation from that rule cannot be altered by policy. The remainder of the normative behavioural rules are to be chosen so as to achieve a vector maximum.

Arriving at this new set of decision rules remains a problem, which Davis and Whinston do not address adequately. To this end, Stiglitz⁸⁶ invokes the idea of deadweight loss as an objective economic criterion to choose between two second best allocations (and their accompanying normative decision rules). Stiglitz argues that second best theory tells us only that we cannot blindly apply the lessons of first best economics. He accepts that choosing a course of action when distortions exist is often difficult. However, he argues that economic theory can tell us when two smaller distortions are preferable to one large one; and when it is better to have inefficiencies in both consumption and production or in production alone.⁸⁶

Stiglitz⁸⁶ thus proposes that while the informational requirements are extensive, variables such as income, expenditure and price can give some assistance to the decision maker.

THE POSITIVE VERSUS NORMATIVE TRADE-OFF IN PRACTICE

Lipsey and Lancaster propose that there is no single necessary, sufficient condition for welfare improvement.^{46,91} Partial equilibrium analysis can identify a

number of second best options but, according to Lipsey and Lancaster, economic theory can do little more than shift society from one second best option to another. The negative corollary to the “General Theory of the Second Best” specifically states that there is no *a priori* way to judge between various situations in which some of the Pareto optimum conditions are fulfilled while others are not.⁴⁶ If partial equilibrium analysis can present us with a number of second best options yet cannot assist us in choosing between these options, how then do we decide on the direction that policy recommendations should take?

If we rely on government officials to make policy decisions, what factors ensure that they act in the general interest? Faith and Thompson^{22,22} and Brown^{11,11} believe that the theory of the second best may assist in explaining observed government policy. If there is no other (positive) criterion for selecting one second best option over another, they may be likely to follow their own self-interest as a policy maker, politician or bureaucrat.^{11,22,91} However, in a truly democratic environment, the risk of being voted out of power is a credible threat. As such, a politician may (out of a sense of self-preservation) choose to act in the interests of his or her constituency. In a less democratic world the outcome may be driven by a desire for self-enrichment on the part of the decision maker, unchecked by any credible threat of sanction. The decision maker would be maximising his or her own individual utility function rather than the social welfare function of the community they serve.

This suggests that there may be ways to contain the extent of self-interested behaviour among those planning and implementing health policy. Social and political institutions could potentially be structured to provide a credible threat of sanction. In addition, income and expenditure data (for example)⁸⁶ could provide a transparent and positive measurement framework for the evaluation of normative rules. A more detailed discussion of this process falls outside the

ambit of this paper but the interested reader is referred most particularly to Stiglitz, (2000).⁸⁶

SOME FINAL THOUGHTS ON THE THEORY OF THE SECOND BEST

Policy design within the world of the second best is a more complicated task than that in the clearly defined Pareto optimal world.⁸⁶ However, partial equilibrium analysis can identify potential second best outcomes,^{46;91} and knowledge of the relevant social welfare function(s) can be gleaned from observable cues such as income data, expenditure data, pricing data and other behavioural measures (including survey data and political polls in some cases).^{86;91} This knowledge of the options, along with society's priorities, can be combined with an analysis of the likely deadweight loss^{86;91} to arrive at an efficient and equitable outcome. Whether or not that allocation is optimal, remains a matter of subjective opinion.^{46;91}

Thus the challenge to cost minimisation; cost benefit; cost utility and cost effectiveness analysis, is to optimise in a first-best world and satisfice in the world of the second best. Let us now consider how effectively each method is likely to meet that challenge in a developing country context.

*Chapter Three: Cost Minimisation – A Rarely Used
Tool*

COST MINIMISATION AND OUTPUT MAXIMISATION

The basic premise of cost minimisation (or its converse, output maximisation) is that, given two health interventions with equal efficacy, the lower cost intervention should be chosen^{48,82}. For example, take the hypothetical case of skin cancer (or melanoma) where a cream or very minor surgery are equally effective at removing the affected cells. If the cream costs less than the surgical procedure, then cost minimisation would advocate the cream. As health is kept constant and only spending is minimised, this has little to do with allocative efficiency in a strict sense. For the purposes of this discussion, allocative efficiency is defined as the allocation of resources across different health interventions so as to maximise the health of a given population^{60,91}. However, if one considers that selecting a lower cost intervention frees up resources for other interventions, cost minimisation does have allocative repercussions in the secondary analysis. When moving from an effective medicine of cost “x” to an equally effective intervention of cost “x - a” (where “a” is a positive integer) a Pareto improvement is realised.

COST MINIMISATION APPLIED IN THE DEVELOPING COUNTRY
CONTEXT

Few traditional cost minimisation (CM) studies have been conducted as health interventions rarely have equal efficacy. However, a recent paper by Mahmud Khan *et al*⁸ uses a more advanced application of the cost minimisation methodology to plan the optimal geographic distribution of maternal care

facilities in Bangladesh. This appears to be the most sophisticated CM analysis recently conducted in a developing country context, hence its use here as an illustration of what CM can (and cannot) measure.

Mahmud *et al*'⁴⁸ model attempts only to reduce the high maternal mortality rates (MMR) in Bangladesh and takes no account of maternal morbidity. The hypothesis is that the availability of emergency obstetric care (EOC) is the most effective strategy to reduce maternal mortality and that the provision of EOC services necessitates the construction of a number of health care facilities.

The cost function being minimised is a social cost function that consists of two parts;

1. The cost of delivery from the provider's perspective (both fixed and variable); and
2. The cost to the household.

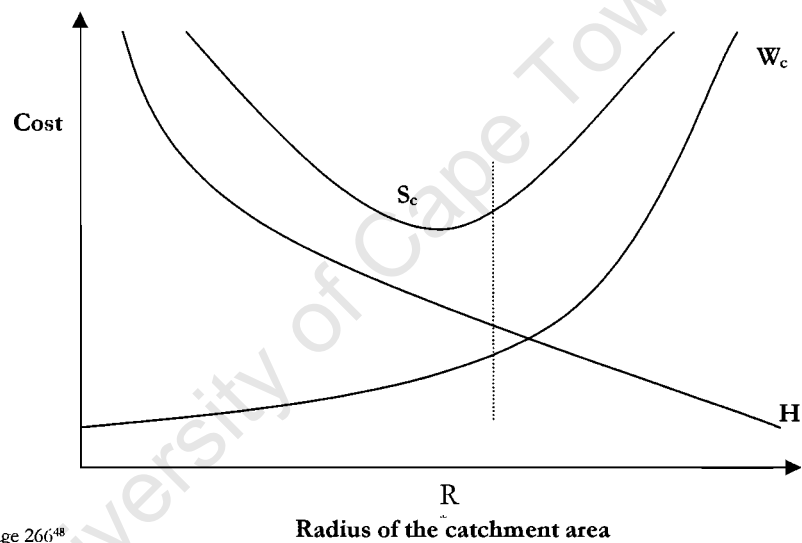
The cost to the household also consists of two components, namely;

1. The cost of accessing the facility (such as transport costs and the opportunity cost of time); and
2. The cost of a lack of access. It is assumed that a lack of access will result in avoidable deaths in some proportion of the cases and this loss of life is given an economic value.

Furthermore, the model assumes that EOC services are provided free of charge to the patient and that the distance to a health centre is the primary factor determining health care utilisation. It is suggested that the distance-decay factor utilised goes some way towards capturing the many social, cultural and economic factors affecting the health care seeking behaviour of women. Although this point is arguable, the simpler structure of the model does have merit.

The costs to the women (W_c) are assumed to be an increasing function of the radius around the health care facility (i.e. the further they are likely to be situated from the facility, the greater the cost). Conversely, the health costs of the national provider (H_c) are assumed to be a decreasing function of the radius around the health centre because the larger the radius served by anyone centre, the fewer health centres are needed to cover the total geographic area of Bangladesh. Conceptually then, the problem looks as follows where S_c represents the social costs calculated by adding W_c and H_c :

Figure One



Source: page 266⁴⁸

The health centre costs (or the provider costs) were measured using the following fixed cost specification:

$$C_f = \text{annualised cost of construction of the facility} + \text{annualised cost of EOC equipment} + \text{maintenance costs} + \text{personnel costs}^\ddagger$$

[‡] It was assumed that personnel costs did not vary according to the number of medical cases because a certain minimum staff complement would need to be retained and paid, regardless of how many cases they treated

The variable costs to the provider were assumed to be the costs of managing a case at the health centre. As the nature of pregnancy complications varies, so do their associated costs. As such, the average variable cost was derived by calculating a weighted average of the costs of treating various categories of complications.

The costs to the household were slightly more complex to calculate. *Table One* lists the parameter values used to estimate the cost to the household of accessing the EOC facility.

Table One

PARAMETER	VALUE USED	SOURCE OF THE DATA
Fraction of pregnancies at risk (α)	0.20	UNICEF
Probability that a high risk patient will use the facility (δ)	0.70	Tertiary hospital data
Fixed cost of transportation (a)	Tk. 4.00	Qualitative survey
Variable cost of transportation (b)	Tk. 2.00	Qualitative survey
Wage rate per hour (w)	Tk. 10	BBS Yearbook, 1997
Hours needed to travel 1km (h)	0.5 hours	Qualitative survey
Death rate among high risk patients if EOC used (λ_H^U)	1.5%	ICDDR,B
Death rate among high risk patients if EOC <u>not</u> used (λ_H^N)	3.0%	Assumed value
Death rate among <u>low</u> risk patients if EOC used (λ_L^U)	0.025%	Assumed value
Death rate among low risk patients if EOC <u>not</u> used (λ_L^N)	0.05%	Derived parameter
GNP per capita in Bangladesh	\$220	World Bank Report, 1996
Years of life lost due to pregnancy related deaths	35 years	BBS Year Book, 1996
Construction cost per square metre	Tk. 665	Ministry of Works (GOB)
Discount rate	5%	
Utilization function	$f(r) = 0.2 - 0.008r$	

Source: page 269⁴⁸

As mentioned before, the model also considers the cost of a lack of access to EOC facilities. To revisit this point; it is assumed that lack of access will result in avoidable deaths in some proportion of the cases and this loss of life is given an economic value. The economic value of life lost is calculated in this model by assuming that the GNP *per capita* measures the economic benefit of each year of life. As such, the economic value of life is represented by the discounted future lifetime income of the individual if premature death were averted.

In short then, the cost of maternal mortality was calculated as follows:

- Each death was assumed to result in a loss of income of US\$220 per year.
- It was assumed that the average age at which maternal mortality occurs is 30 and that life expectancy is 65, resulting in a loss of 35 years of life.
- The future income stream was discounted at a rate of 5% per annum.

Table Two shows how the findings of the model were presented for one area in Bangladesh. For that area the optimal radius was found to be 6 km however, the model was able to inform national planning as follows:

*“For most districts in Bangladesh, the minimum average cost per case occurs when the catchment area radius of a health facility providing EOC is about 10km...if the optimum radius is 10km, the [optimal] catchment area of a centre in the district should be πr^2 or, 314 km² (3.14*10²). The catchment area is then used to calculate the total number of facilities needed to cover the whole population in a district. Adding the number of facilities needed in all the 20 districts gives us the total number of EOC facilities recommended for Bangladesh.” page 270⁴⁸*

Table Two

Radius (R)	Utilization (U) Per Centre Per Month (No. Of Women)	Travel And Opportunity Costs Per Health Centre	Medical Care Costs Per Case Managed	Annualised Capital Cost Per Centre	Cost Due To Death Associated With A Centre	Total Average Cost Per User
2	5	198	1111	60539	23156	17 890
4	19	1207	1111	60539	93611	9 288
6	39	3573	1111	75673	212843	8 600
8	64	7606	1111	105942	382331	8 859
10	92	13424	1111	121077	603 553	9 133
12	123	20937	1111	181615	877 988	9 896
14	154	29856	1111	196750	1 207 116	10 421
16	184	39690	1111	227019	1 592 414	11 215
18	198	49724	1111	251235	2 035 360	12 910
20	223	59116	1111	272424	2 537 434	13 976
22	243	66709	1111	287558	3 100 115	15 326

Source: page 27¹⁸

SHORTCOMINGS OF THE COST MINIMISATION METHODOLOGY

Some obvious criticisms of the cost minimisation model as it is applied in this example include the use of a human capital approach to the valuation of life (the problems with this approach will be discussed in more detail in Chapter Six). The authors do point out however, that the model can accommodate alternative measures of human value, as long as an economic value of life of some kind is included.^{48,91}

A second criticism arises from the life expectancy assumption, which appears to be unsubstantiated by any demographic foundation (again, the merits and

demerits of various life expectancy approaches will be discussed in more detail in the next chapter). The authors do not explicitly address this problem.

Finally, the authors acknowledge that the model fails to take morbidity factors into account. They acknowledge that this will affect the findings of the model by reducing the optimal size of the radius (and hence the optimal size of the catchment area) to some degree.

A less obvious problem with this model stems from its first-principles (welfare economic) foundation, or its lack thereof. The model is relevant to the context of maternal mortality only. However, building new facilities may well require an increase in the health budget or, at the very least, a restructuring of the health budget. Where will the additional funds be taken from and at what “cost”? These broader allocative implications are not accounted for in the model.

A FINAL WORD ON COST MINIMISATION

Despite these criticisms, Mahmud Kahn *et al*⁴⁸ prove that the cost minimisation methodology can be useful for at least loosely informing policy planning, although it could be argued that cost effectiveness analysis might have informed this decision in much the same way[§]. Mahmud Khan *et al* have ensured that their model can be applied in contexts outside of Bangladesh and, with some adjustment, it is possible that it could be used to inform other (similar) problems while taking into account the criticisms levelled above.

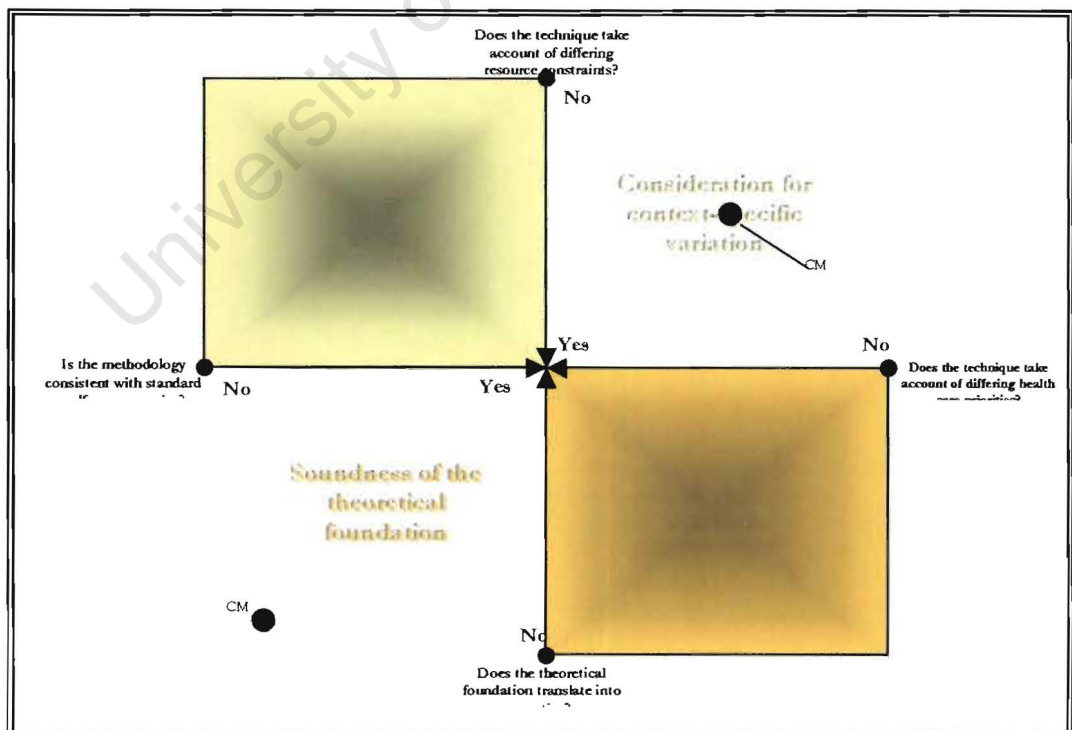
The shortcomings detailed above place cost minimisation in the following positions on the assessment matrix detailed earlier (see *Figure Three*). Cost minimisation fares moderately well in terms of context-specific variation as it has been shown to have some value in the developing country context. However, as

[§] See Chapter Four for more discussion on cost effectiveness analysis (CEA).

far as context-specific variation is concerned, cost minimisation fails to account for morbidity factors and to explicitly consider possible differences in health priorities. Maternal mortality is assumed to be a significant health problem in Bangladesh because the maternal mortality ratio fares unfavourably with more developed countries. However, this does not necessarily lead to the conclusion that maternal mortality should be Bangladesh's most urgent health care priority. The cost minimisation exercise detailed above sheds no light on this issue.

As discussed, cost minimisation fares relatively poorly in terms of its theoretical foundation, primarily because it takes little account of the allocative implications of health spending. It must be emphasised that the position of a methodology on the comparative matrix is the result of a qualitative assessment only which is, of necessity, a conceptual and hence a subjective framework. The specific co-ordinates on the matrix are somewhat arbitrary.

Figure Three: Positioning CM within the assessment matrix



Chapter Four: Cost Effectiveness Analysis - An Inadequate Reflection of Social Welfare Ideology

INTRODUCTION

Although for reasons of clarity, cost effectiveness is clearly differentiated from cost benefit and cost utility analysis in this paper, there is some overlap in their usage in practice. Cost utility analysis is often considered to be a form of cost effectiveness analysis that enables one to arrive at a cost per util measure rather than a cost per unit of health output (e.g. life year saved, cases cured etc). However, for the purposes of this discussion, a strict analytical separation will be maintained.

Typically, cost effectiveness analysis (CEA) describes a health intervention in terms of marginal cost over marginal benefit. In practical terms this amounts to the measurement of incremental costs over incremental health benefits.^{25-27,91} These ratios of relative cost effectiveness are then used as a way to rank the technical efficiency of alternative medical interventions. Given that different interventions often have very different health outcomes, this generalisation is usually accomplished by reducing the output to a common denominator such as life years saved.^{**42,91} Currie *et al*^{15,91} cite the CEA as the promoted method for identifying:

- The least cost way of achieving a given output, or
- Whether the same level of output could be achieved with fewer inputs; or
- The optimal way to spend a given budget.

** Cost utility analysis adjusts these life years saved. The merits of this adjustment will be discussed in more detail in Chapter Five.

THE APPARENT SOCIAL WELFARE FOUNDATION

Despite the frequent use of CEA for economic evaluations of health interventions,^{9;34;38;41;42;70;71;81;106} there does not appear to be any justification for the methodology on the grounds of first principles welfare economics.^{15;25;91} Both Garber and Phelps²⁶ and Currie *et al*^{15;91} contest the use of cost effectiveness analysis on social welfare grounds. The core of each of their arguments is that the first principles foundation of CEA is sound when the analysis is taking place on the individual level but not when it is taking place at the societal level. On an individual level there are few allocative implications of health spending. Those allocative implications that exist can be captured by the individual's utility function. However, on the societal level, inter- and intra-budgetary health spending has allocative implications poorly captured by the CEA methodology. This will be illustrated shortly using an example.^{26;91}

Why is the CEA used so often if it has such dubious theoretical foundations? Garber and Phelps attribute the popularity of the CEA to the familiar ring of the CEA problem:

“ ... (minimising the cost of producing a given level of health, or correspondingly, maximising the achievable level of health for a given budget) sounds like a familiar economic problem, and for the most part, practitioners have assumed that CE analysis could be a tool for utility maximisation...”

page 1²⁶

Currie *et al* take a similar line, pointing out that the theoretical foundation of CEA lies in generally accepted production theory where the point of tangency between the isocost and the isoquant denotes the most efficient point of production.^{15;91}

Despite this general application of the CEA methodology in the health care context, Garber and Phelps^{26,91} point to three problems raised by the literature. The first is whether CEA should include future medical costs incurred solely because of the prolonged period of life. The second is whether the use of life years as a health outcome results in discrimination against older persons. The third relates to the identification of an accepted cost effective cut-off ratio; how effective is cost effective?

“Although one might conclude that equalization of cost-effectiveness ratios at the margin is necessary for Pareto optimality, there is still the question of the proper level” page 2²⁶

The problem of age discrimination is dealt with in considerable detail in Chapter Five. However, the simple Garber and Phelps^{26,91} model of cost effectiveness explained below, illustrates many of the salient points regarding the welfare economic foundations of CEA. The model also addresses the derivation of an appropriate CEA cut-off ratio.

THE INDIVIDUAL CASE: A SOUND “FIRST PRINCIPLES” FOUNDATION

Garber and Phelps^{26,91} set up the following simple three period model in which the individual has a Von Neumann-Morgenstern expected utility function. For each period, utility is a function of income net of medical care expenditures and all individuals are alive in period 1. The only relevant choice variable is medical care expenditure in period 1 as medical care expenditure in period 2 is assumed to be independent of expenditure in period 1.

$$E(U) = U_1(Y_1 - C_1) + P_2(C_1)U_2(Y_2 - C_2) + P_2(C_1)P_3(C_2)U_3(Y_3)$$

Equation 1

- Y_i : Individual income in period i
- C_i : Medical care expenditures in period i
- P_i : The probability of surviving from one period to the next (assuming $P_1 = 1$)
- U_i : Individual utility in each period

While equation one will give us an individual's expected utility given a choice of medical care "consumption" in period one, it is also possible to use this framework to solve for the individual's optimal investment in C_2 by differentiating consumption in C_1 by the probability of surviving to P_2 given that the effectiveness of the intervention in this model is the increase in the probability of surviving from one period to the next. Garber and Phelps structure this optimisation problem as follows:

$$\frac{dC_1}{dP_2} = \frac{U_2(Y_2 - C_2^*) + P_3^*U_3(Y_3)}{U_1'}$$

Equation 2

They conclude from the above optimal incremental CE ratio, that an optimum is reached when the CE ratio equals the sum of future expected utility (normalised by the marginal utility of income in period 1 as denoted by U_1'). In the individual context, this does go some way towards defining an optimal CE ratio for medical resource decisions by using the principles of expected utility maximisation principles.

Equation two fails to address the question of unrelated future medical costs i.e. costs that have no relevance to the current treatment or illness but which are

incurred because the individual lives longer. Garber and Phelps^{26,91} attempt to redress this by defining a total lifetime medical cost function as:

$$C^{t\alpha} = C_1 + P_2(C_1) \bullet C_2$$
Equation 3

From this definition one can see fairly easily that:

$$\frac{dC^{t\alpha}}{dP_2} = \frac{dC_1}{dP_2} + C_2$$
Equation 4

Thus the optimal cut-off is the same as before with the addition of period-2 medical expenditure (or costs). This returns us to the issue of normative decision rules discussed earlier. Garber and Phelps recommend that the analyst practises consistency by selecting the CE cut-off that corresponds to the chosen cost accounting method. However, they don't go into detail about the possible choice of cost accounting method or how the CE ratio should correspond.

Garber and Phelps also recommend that the dictates of parsimony and the difficulty in measuring future health expenditures, could serve to justify omitting them. However this very difficulty in measurement makes it impossible to measure accurately the impact of the omission. There does not appear to be any published literature that deals with this issue in any detail and this measurement challenge may well be the reason for the dearth.

If one is drawing on the first principles foundation discussed earlier, analysts should be seeking to equate the marginal benefits and costs of all inputs. This is exactly what Garber and Phelps^{26,91} recommend that: *“one should adjust the intensity of all medical interventions so that they have a common CE ratio”* (page 9). This consistent selection criterion for health interventions is seen as a particular strength of the CEA methodology.

THE SOCIETAL CASE: A “FIRST PRINCIPLES” IMPOSTER

“The variability of the optimal CE ratio across persons leads to a fundamental tension in using it to guide the allocation of health care resources: insurers and policy makers may wish to equate CE ratios across interventions and across populations, yet the members of the population may have very different optimal CE ratios....CE analysis applied at the population level may give the most efficient egalitarian distribution of health resources, but it is not likely to be Pareto optimal.” page 29²⁶

Garber and Phelps reach a conclusion which has considerable implications for the planning of public health provision; namely, that the optimal CE ratio is most sensitive to two individual characteristics: income and risk aversion.

As these characteristics differ considerably from individual to individual – how is the government planner expected to optimise health allocations in a culturally, economically and socially heterogeneous society? In a country where the gap between the very wealthy and the very poor is wide, can one simply take a mean income level and some mean measure of risk aversion and apply it to the total populace?

Currie *et al*^{15,91} propose that the societal approach to cost effectiveness contradicts the textbook idea of CEA. They use an example that takes into account the patients’ costs of an intervention e.g. transport to the clinic. They show that this can result in the selection of an intervention that is most cost effective from a patient perspective. However if an alternate intervention is more cost effective from a pure provider perspective, then one must consider the opportunity cost of those provider funds “wasted” on the less cost-effective intervention. Those funds could (in theory) have been allocated to other patients needing other forms of care.

Cost effective analysis is most pertinent when alternate programme designs are being compared. In fact the cost effectiveness measure was developed to enable policy makers to select a bundle of cost effective health interventions. This is done by selecting those ratios with the lowest cost per life year saved.^{60,91} However, as the quote below suggests, by simply comparing cost effectiveness ratios no credence is given to affordability or absolute effectiveness.

“It is not possible with cost-effectiveness analysis to determine whether a treatment is profitable for society or not. Cost-effectiveness analysis provides a direct answer only if the alternative with the best effect has the lowest cost effectiveness ratio. If one alternative has a higher cost-effectiveness ratio than another but offers better effects, the choice between these alternatives cannot be made simply on the basis of cost-effectiveness analysis, but some decision criterion has to be introduced.” pg 6⁴⁰

A recent paper measuring the cost effectiveness of highly active anti-retroviral therapy (HAART) for HIV positive patients in South Africa estimated the cost per life year saved to be between R5,600 and R5,700 (approximately).^{10,42} This is more cost effective than kidney dialysis, chemotherapy, coronary by-pass surgery and a number of other interventions offered by the South African public health service. Despite this, HAART is not currently offered in the public health sector. One of the factors that may well be giving policy-makers pause is the prevalence rate of the disease in South Africa. Far more patients are likely to apply for HAART than for chemotherapy potentially resulting in a higher total cost of delivery⁶. This higher total cost might have significant allocative implications, which are not addressed in the cost effectiveness framework despite their particular importance in the economically disparate society of South Africa.

⁶ Boule *et al*⁹ estimated that the total program costs for 2007 would be around R300 million. This assumed approximately 84,000 people or 10% of all Stage Four AIDS patients on treatment.

Currie *et al* argue that, “there is no role for a societal approach to CEA” (page 1)^{15,91} because the societal approach raises allocative issues which the CEA methodology is ill-equipped to measure. Boyle *et al* (as referenced in Currie *et al*) make the point that a CEA which takes a societal approach, should take into account all the associated costs and benefits of an intervention, regardless of where their benefit or burden may fall. Currie *et al* take this definition to include costs and benefits that may fall outside the context of health services (and the health budget) and they point to the significant distributional issues that this broader perspective necessarily raises. Of course, these shortcomings assume that the CEA is applied mechanically without any political and/or social consideration of the use of the findings or indeed, the initial definition of the problem and the budget available for its solution.

Currie *et al*¹⁵ suggest that cost effectiveness calculations are often done when cost benefit calculations are a more appropriate cost measure. Cost benefit analysis is a technique employed in identifying, quantifying and valuing – in a common yardstick such as money – all important costs and consequences to society of any potential or actual change in resource use allocation in the economy. As such, cost benefit analysis should (in theory) enable the policy maker to take into account the value of the forgone treatments.

AN ADDITIONAL CRITICISM OF THE CEA METHODOLOGY

Garber and Phelps²⁶ cite a wealth of literature that identifies the interaction of individual characteristics, and most particularly the interaction of age and risk aversion:

“...at higher degrees of risk aversion, the optimal CE ratio shifts more with age, and as people become less risk averse, age has a diminishing (and finally nearly zero) effect on the optimal CE ratio” page 26-27²⁶.

Murray^{58,91} makes the point that adding years to life does not have the same meaning at all stages of life. For example, adding years to life when one is a child means adding active years of learning and growth, however, as Murray points out, adding years to the latter stages of life may add very little in terms of life quality. It is this difference in life quality (which is not only determined by age!) that the conventional CEA fails to capture. Garber and Phelps see this as a key failing and they argue that by omitting the effects of an intervention on future quality of life, particular treatments are likely to be undervalued (here they cite the example of rehabilitative care).

This thinking has led to the practise of quality adjusting life years to capture the burden of disease when that burden does not necessarily result in death. The critical assessment of this practise will be done as part of a broader discussion on Cost Utility Analysis (CUA) in the next Chapter.

A FINAL WORD ON COST EFFECTIVENESS ANALYSIS

“Resource allocation decisions can never be shaped by the mechanical ranking of cost –effectiveness ratios. Ratios provide information about one type of “value”, health benefit per dollar spent. But other types of values of society, including considerations of distributive justice and fairness...require that CEA be viewed as an informer of decision making rather than a decision maker per se.” [Gold et al (1996)]^{‡‡}

‡ As quoted in Currie *et al*¹⁵, pg 14

Conclusions, and hence policy recommendations, cannot necessarily be made by a simple comparison of cost-effectiveness ratios.^{15,21,21,91} Cost effectiveness measures merely succeed in giving a ranking of various interventions in terms of efficiency, usually measured in terms of cost per life year or cost per health outcome. The most cost effective option is not necessarily the cheapest option in terms of total costs. This variation in total costs raises a number of policy concerns. Firstly, cost effectiveness measures say nothing about whether the health budget can accommodate an intervention. Secondly, it cannot be certain that allocating the additional funds to an intervention that is more cost effective (but has a higher total cost) is in fact the best use of those funds. The opportunity cost of the additional funding has not been accounted for and until it is, any concrete policy recommendations would be ill advised in any setting. What this implies of course, is that many of these shortcomings could be overcome if the CEA were conducted in a dynamic socio-political environment in which issues such as budget allocations could be considered without the undue influence of special interest groups (for example). This may not always be the case in the developing country context – hence the critical treatment of the CEA in the context of this discussion.

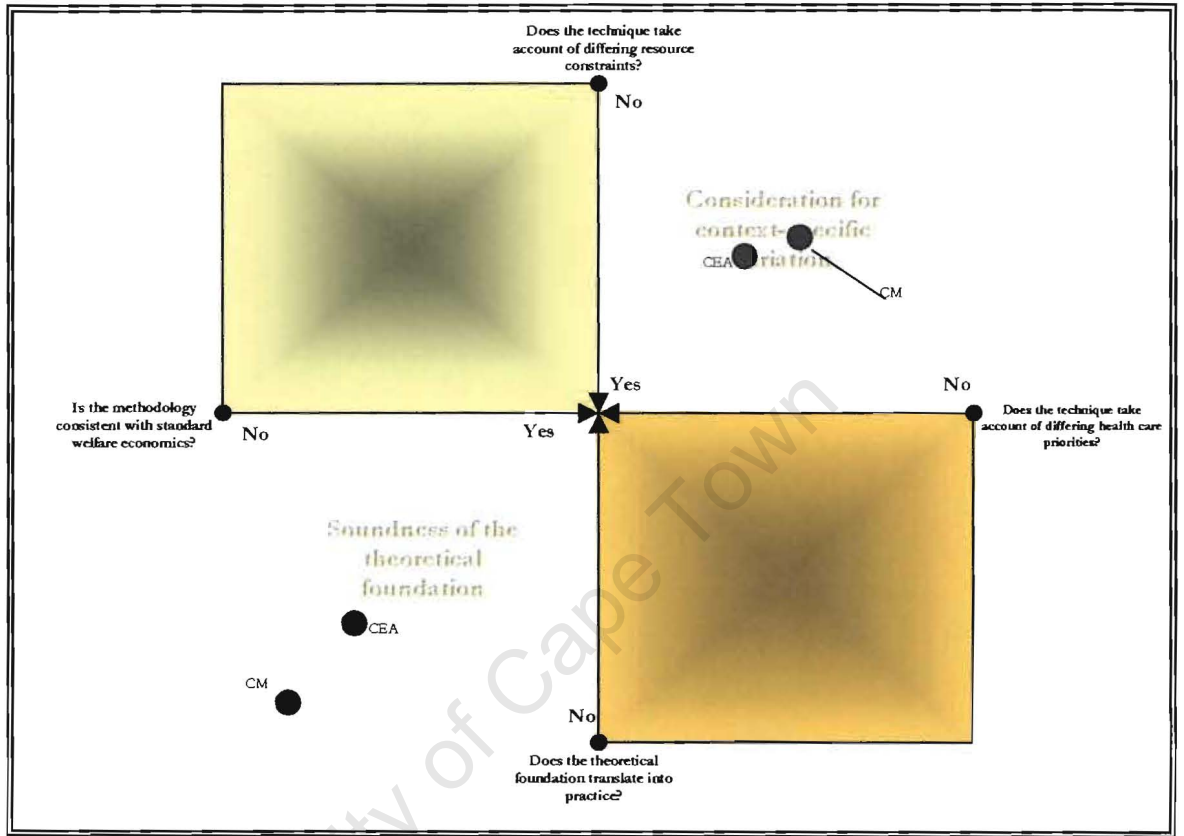
In the language of the formal assessment matrix, CEA fares better than cost minimisation (CM) in terms of both theoretical foundations and context-specific variation. In terms of context specific variation, Boule *et al*¹⁰ have shown that CEA can inform health policy by contextualising the cost of new interventions (such as HAART) along side existing health care interventions (such as kidney dialysis). However, CEA is still found wanting in many respects as it serves only to rank health interventions without advising a cut-off point or an optimal program size for each intervention. Furthermore CEA says nothing about the interaction of various interventions (such as HAART and HIV prevention strategies.)^{49,54} These “interactions” or synergies can offer significant positive

externalities in resource constrained environments. CEA does not assist with the setting of health priorities on any basis other than direct cost.

On a theoretic basis, both Currie *et al*¹⁵ and Garber and Phelps²⁶ argue that CEA has sound economic foundations at the individual level but poor foundations at the societal level.

University of Cape Town

Figure Four: Positioning CEA within the assessment matrix



*Chapter Five: Cost Utility Analysis – Is There a
Way to Reflect Accurately the Utility of Health
Care?*

WHY DO WE NEED A COMPOSITE INDICATOR OF BURDEN OF
DISEASE?

The term “burden of disease” (BoD) commonly refers to both morbidity and mortality, where morbidity is the variation from perfect health experienced by any ill person. In composite burden of disease measures, disability weights attempt to capture degrees of suffering or incapacity associated with different non-fatal conditions (or fatal conditions which are preceded by a loss of physical function.)³⁹ For example, the Quality Adjusted Life Year (QALY) sets perfect health equal to one and death equal to zero.^{26,91} The burden of disease in this composite measure thus varies between one and zero suggesting that a QALY is a “good” to be maximised. Conversely, the Disability Adjusted Life Year (DALY) sets perfect health equal to zero and death equal to one. This results in similar variation between zero and one, but the DALY should be considered a “bad” to be minimised.^{3,4,91}

The arguments against measures of the burden of disease are varied and include concerns about the accurate measurement of the burden,^{37;56;91;98;99} the definition of the burden across settings,^{23;37;56;58;61;91;98;99} the incompleteness of the concept^{23;37;56;58;61;91;98;99} and the inconsistency of the concept with social welfare economics in its violation of (or disregard for) Arrow’s impossibility theorem^{§§}; i.e. that one cannot sum the individual utilities to arrive at a social utility function.^{***}

§§ This criticism does not appear in the current literature but is valid none the less.

*** BoD is seldom measured at the individual level. It is usually generalised to all sufferers of a particular illness.

Murray^{58;59;61;91} reduces the controversy around BoD indicators to the basic tenants of the implicit versus explicit rationing dilemma^{50;51}. He lists four objectives that any BoD indicator should aim to satisfy^{58;59;61;91} and cites the first two objectives as an indication that a BoD measure is in fact a rationing aid:

1. To aid in setting health service priorities
2. To aid in setting health research priorities
3. To aid in identifying disadvantaged groups and targeting health interventions
4. To provide a comparable measure of output across the forms and stages of an intervention

It is Murray's^{58;91} belief that except for the case of completely free markets, the allocation of resources between programmes defines a set of relative weights for different health outcomes which are likely to be assigned in a less than transparent manner.

"If the process of choosing relative weights of different types of health outcomes is left entirely to the political or bureaucratic process there is a high probability that similar health outcomes may be weighted inconsistently... there may be no open discussion or debate on key value choices or differential weightings." pgs 429-430⁵⁸

The benefit of a measure such as the DALY assumes then, that the political process of priority setting is closed and that a burden of disease indicator would somehow allow a more open and explicit form of rationing. While it is true that a substantial amount of digging through the literature will reward the reader with the components of the DALY and the assumptions inherent in its design, it is not true that this information is generally available or accessible, let alone easily understood. The arguments for and against burden of disease indicators contain some elements of the implicit versus explicit rationing argument, but the DALY

may not be the explicit tool that Murray supposes^{†††}. Similarly, institutional (or political) rationing mechanisms may not be the obscure (or entirely implicit) tool that Murray criticises. Recommended protocols such as the “Standard Treatment Guidelines and Essential Drugs Lists” explicitly ration doctor’s treatment choices in South Africa.^{42,85}

While the implicit versus explicit rationing debate is one component of the BoD debate – it is certainly not the only component of that debate. To reduce the controversy to the basis for rationing is to obscure problems that arise from Murray’s latter two objectives of the BoD indicator. These last two objectives speak of comparability. For an intervention (or its output) to be comparable across different contexts it must first be reduced to a common or generic basis. Proponents of the DALY^{58-61,91} argue that this is a key benefit, utility is common to all economic agents and as such a cost per DALY measure can be compared across social groups within a country, as well as between countries and continents. A concern given less attention in the literature is that by reducing the BoD to satisfy this generic requirement, assumptions and value judgements must be made which may not in fact apply across all of these settings.

Mooney *et al*^{56,91} express concern about the desire to generate a uniform measure of the burden of disease⁶¹. They make the point that the burden of disease is likely to vary between communities and within communities. A paraplegic working in an office environment is likely to be less hindered by disability than a subsistence farmer in a rural area who needs to be able bodied in order to grow enough food for the family. Similarly, for a monthly wage earner who is paid for “sick-leave”, the burden of disease is likely to be less than for the hourly wage earner who is not paid if they do not turn up.

††† The following empirical example will illustrate how changes in the assumptions of the DALY can impact on the resulting policy recommendations. The DALY does not appear to be a standard rule of measurement but, instead, a rather variable rule of measurement.

WHAT IS A DALY^{##}?

Traditional indicators of health such as life expectancy, fertility and infant, child, maternal and adult mortality, have largely been replaced in the health economic literature by composite measures of both mortality and morbidity (such as DALYs and QALYs). Conventional mortality figures fail to capture the effects of sickness and disability (morbidity) which do not result in death, and statistics on disease incidence can be rough at best. Furthermore, morbidity statistics are likely to be particularly poor in less developed countries where reporting mechanisms may not be functioning optimally and where a lack of access to medical care may prevent any formal recording of morbidity incidence. Similarly, as morbidity statistics are generally collected at the point of treatment, the few morbidity statistics that do exist tend to be biased towards the wealthier segment of the population, which can afford to seek medical care more frequently than poorer segments.¹⁰⁴ The disability adjust life year (DALY) is thus a quantitative indicator of burden of disease (BoD) that reflects the total amount of healthy life lost, whether to morbidity or premature mortality.³⁹

As was the case with CEA, cost utility analysis (CUA) and DALYs were designed to assist in setting national and global health priorities. The principle suggestion of the 1993 World Development Report¹⁰⁰ was to use CUA to select a package of essential health services. The line between the essential and discretionary interventions was to be drawn on the basis of their cost effectiveness³⁹.

This suggests that the DALY (or the process of minimising DALYs) can assist in optimising allocative efficiency. However, the extent to which one accepts this premise should be tempered by an understanding of how effectively the DALY reflects the level of morbidity in a society. It should also be tempered by the fact

^{##} This paper focuses on the DALY as this is the composite measure of BoD recommended by the World Bank and the World Health Organisation for use in the developing country context.¹⁰⁰

that not all public spending allocations will have direct morbidity implications. For example, the link between education and morbidity is tenuous at best – surely proponents of the DALY would not suggest that the entire education budget be reallocated to health care in order to minimise the level of morbidity in a population? The allocative strengths of the DALY should therefore be treated with caution.

EXAMPLES OF THE LEADING CAUSES OF MORTALITY AND MORBIDITY UNDERLYING DALYS AROUND THE WORLD

*Table One*¹⁰³ over the page shows that HIV is the leading cause of mortality and morbidity underlying DALYs amongst 15-44 year olds around the world.

Table One: Leading causes of disability-adjusted life years

Figure 2.2 Leading causes of disability-adjusted life years (DALYs), in all ages and in 15–44-year-olds, by sex, estimates for 2000^a

Both sexes, all ages		% total	Males, all ages		% total	Females, all ages		% total
1	Lower respiratory infections	6.4	1	Perinatal conditions	6.4	1	HIV/AIDS	6.5
2	Perinatal conditions	6.2	2	Lower respiratory infections	6.4	2	Lower respiratory infections	6.4
3	HIV/AIDS	6.1	3	HIV/AIDS	5.8	3	Perinatal conditions	6.0
4	Unipolar depressive disorders	4.4	4	Diarrhoeal diseases	4.2	4	Unipolar depressive disorders	5.5
5	Diarrhoeal diseases	4.2	5	Ischaemic heart disease	4.2	5	Diarrhoeal diseases	4.2
6	Ischaemic heart disease	3.8	6	Road traffic accidents	4.0	6	Ischaemic heart disease	3.3
7	Cerebrovascular disease	3.1	7	Unipolar depressive disorders	3.4	7	Cerebrovascular disease	3.2
8	Road traffic accidents	2.8	8	Cerebrovascular disease	3.0	8	Malaria	3.0
9	Malaria	2.7	9	Tuberculosis	2.9	9	Congenital abnormalities	2.2
10	Tuberculosis	2.4	10	Malaria	2.3	10	Chronic obstructive pulmonary disease	2.1
11	Chronic obstructive pulmonary disease	2.3	11	Chronic obstructive pulmonary disease	2.4	11	Iron deficiency anaemia	2.1
12	Congenital abnormalities	2.2	12	Congenital abnormalities	2.2	12	Tuberculosis	2.0
13	Measles	1.9	13	Alcohol use disorders	2.1	13	Measles	2.0
14	Iron-deficiency anaemia	1.8	14	Measles	1.8	14	Hearing loss, adult onset	1.7
15	Hearing loss, adult onset	1.7	15	Hearing loss, adult onset	1.8	15	Road traffic accidents	1.5
16	Falls	1.3	16	Violence	1.6	16	Osteoarthritis	1.4
17	Self-inflicted injuries	1.3	17	Iron-deficiency anaemia	1.5	17	Protein-energy malnutrition	1.2
18	Alcohol use disorders	1.3	18	Falls	1.5	18	Self-inflicted injuries	1.1
19	Protein-energy malnutrition	1.1	19	Self-inflicted injuries	1.5	19	Diabetes mellitus	1.1
20	Osteoarthritis	1.1	20	Cirrhosis of the liver	1.4	20	Falls	1.1

Both sexes, 15–44-year-olds		% total	Males, 15–44-year-olds		% total	Females, 15–44-year-olds		% total
1	HIV/AIDS	13.0	1	HIV/AIDS	7.1	1	HIV/AIDS	13.9
2	Unipolar depressive disorders	8.6	2	Road traffic accidents	7.7	2	Unipolar depressive disorders	10.6
3	Road traffic accidents	4.9	3	Unipolar depressive disorders	6.7	3	Tuberculosis	3.2
4	Tuberculosis	3.9	4	Alcohol use disorders	5.1	4	Iron deficiency anaemia	3.2
5	Alcohol use disorders	3.0	5	Tuberculosis	4.5	5	Schizophrenia	2.8
6	Self-inflicted injuries	2.7	6	Violence	3.7	6	Obstructed labour	2.7
7	Iron deficiency anaemia	2.6	7	Self-inflicted injuries	3.0	7	Bipolar affective disorder	2.5
8	Schizophrenia	2.6	8	Schizophrenia	2.5	8	Abortion	2.5
9	Bipolar affective disorder	2.5	9	Bipolar affective disorder	2.4	9	Self-inflicted injuries	2.4
10	Violence	2.3	10	Iron deficiency anaemia	2.1	10	Maternal sepsis	2.1
11	Hearing loss, adult onset	2.0	11	Hearing loss, adult onset	2.0	11	Road traffic accidents	2.0
12	Chronic obstructive pulmonary disease	1.5	12	Ischaemic heart disease	1.9	12	Hearing loss, adult onset	2.0
13	Ischaemic heart disease	1.5	13	War	1.7	13	Chlamydia	1.9
14	Cerebrovascular disease	1.4	14	Falls	1.7	14	Panic disorder	1.6
15	Falls	1.3	15	Cirrhosis of the liver	1.6	15	Chronic obstructive pulmonary disease	1.5
16	Obstructed labour	1.3	16	Drug use disorders	1.6	16	Maternal haemorrhage	1.5
17	Abortion	1.2	17	Cerebrovascular disease	1.5	17	Osteoarthritis	1.4
18	Osteoarthritis	1.2	18	Chronic obstructive pulmonary disease	1.5	18	Cerebrovascular disease	1.3
19	War	1.2	19	Asthma	1.4	19	Migraine	1.2
20	Panic disorder	1.2	20	Drownings	1.1	20	Ischaemic heart disease	1.1

^aNeuropsychiatric conditions and self-inflicted injuries (see Annex table 3) are highlighted.

CONSTRUCTING A DALY

The DALY calculation incorporates five key “social preferences” in an attempt to measure the burden of disease:^{23,39,74,97}

1. The duration of time lost due to premature mortality
2. Disability weights
3. Age Weights
4. Time preferences
5. Health can be summed across individuals

Each of these will now be discussed in more detail before constructing the full formula for the DALY.

1) The duration of time lost due to premature mortality

As both Murray⁶¹ and Molla *et al*⁵⁵ point out, the measurement of health status using time is not a new idea. However, specifically what form this measurement should take seems to be the subject of some dispute. Murray,⁶¹ places the commonly used approaches into four distinct categories:

i. The “potential” years of life lost approach:

$$\sum_{x=0}^{x=L} d_x (L - x) \quad \text{Equation 1}$$

This method is calculated by defining a potential limit to life (L) and calculating the years lost due to age at each death as the limit less the age at death, where d_x in the above formula is the number of deaths at age x. One key failing of this method is that deaths occurring after the imposed age-limit (or limit to life “L”) do not contribute to the estimated burden of disease. Other failing is the fact that the selection of that limit “L” can be largely subjective even if it is informed by available data.⁶¹

ii. **The “period” expected years of life lost approach:**

$$\sum_{x=0}^{x=l} d_x e_x \quad \text{Equation 2}$$

In the formula above, e_x is the expectation of life at each age. This method uses the local life expectancy as the estimate of duration of life lost at each age and may be considered to be more accurate as it estimates the stream of life within a particular context of risk of death from various causes. One problem with this method is that the differing values of life expectancy would lead to the conclusion that the death of a 40 year old woman in Soweto contributes less to the global burden of disease than the death of a 40 year old woman in the London because the expectation of life at age 40 is lower in South Africa than it is in the United Kingdom.⁶¹

iii. **The “cohort” expected years of life lost approach:**

$$\sum_{x=0}^{x=l} d_x e_x^c \quad \text{Equation 3}$$

In the above equation, e_x^c is the life expectancy of a cohort at each age. This cohort life expectancy must be estimated as one cannot know the mortality experience of a cohort in the future. The key drawback of this method is that, like the period approach, cohort life expectancy will differ from community to community, bearing with it an imputed bias against poorer communities.⁶¹

iv. The “standard” expected years of life lost approach:

$$\sum_{x=0}^{x=l} d_x e_x^* \quad \text{Equation 4}$$

This is the approach used for the calculation of DALYs where e_x^* in the above expression is the expectation of life at each age based on some ideal standard.⁶¹ The duration of time lost to death at each age is calculated by deducting the age at death from the ideal life expectancy of the individual. The ideal life expectancy used in the DALY matches the highest internationally observed life expectancy, which is that of Japanese women (82 years). This corresponds closest to Brass’ West model life table, level 26, which has a life expectancy at birth for females of 82.5 and a life expectancy at birth for men of 80 years.³⁹

Murray describes the task of constructing a BoD measure as taking “an *n*-dimensional matrix of health outcomes and collapsing this into a single number” pg 430⁶¹. The subjective element of this process however, lies in deciding which variables to include. The DALY restricts that set of variables to those characteristics which are common to all communities and households, namely age and gender.

“Variables defining subgroups such as income or education, which not all individuals or households can hope to belong to, are expressly excluded from consideration,” pg 431⁷⁴.

Life expectancy in South Africa currently stands at approximately 47.8 years.¹⁰¹ However, while reducing the life expectancy in the DALY to this figure might more accurately reflect the burden of disease in South Africa, it would also bias against treatments in developing countries as they would lose fewer life years due to premature mortality.

Another problem with this calculation is the relationship between BoD and life expectancy itself. The burden of disease, and particularly that of HIV, has played

a key role in reducing the life expectancy of the average South African. This creates something of a multi-collinearity problem as life expectancy is a variable in the BoD composite measure, but the BoD experienced can affect one's life expectancy. If one assumes then that the life expectancy observed in Japan reflects the best health one can hope to enjoy then we do go some way towards removing the effects of health on duration of life. However, the troubling fact is that health is not the only determinant of life expectancy. Life-style, cultural norms, income and other factors result in different life expectancies for different countries and for different social groups within countries.

The trade off then is as follows: to have a measure that can be compared across social groups and across countries but which less than accurately reflects the burden of disease in each of those settings individually; or to have a measure which accurately reflects the burden of disease in homogenous groups, but which cannot be used to compare the effectiveness of measures across social or cultural contexts.

2) *Disability weights*

It is Murray's⁶¹ belief that any health outcome that represents a loss of social welfare should be included in an indicator of BoD, if the society concerned is willing to devote some resources to avert or treat it. Murray clearly points out however, that in mentioning social welfare he is not touting the DALY as the best measure of the health component of social welfare nor arguing that maximising DALYs averted by an intervention amounts to maximising social welfare.

Mooney *et al*^{56,91} also point out that the existing concept of burden of disease (BoD) is incomplete as it only measures the negative impact of illness. For example, it does not measure the positive impact of believing that your country has a comprehensive social health care system (as is the case in Australia). This

positive externality is not captured by current BoD measures. Mooney^{56,91} suggests that were these positive externalities to be captured, more resources might be allocated to the aboriginal people of Australia than are currently allocated. Thus the current concept of “burden” can lead to bias in the allocation of resources.

While disability weights are an interesting measure in their own right, they are highly context specific and sensitive to societal bias.^{65,91} BoD should thus be avoided as a tool for policy setting unless it is supplemented by additional analysis which compensates for these deficits.

3) Age Weights

The age weights are designed to capture the relative importance of healthy life at each stage of life. The weights rise until the age of 25 and then decline slowly thereafter.³⁹ The age weight function is an exponential function of the following form:

$$\text{Age weighting function} = Cxe^{-\beta x}$$

Equation 5

Where:

- C is a constant = 0.16243
- β is a constant = 0.04
- x = age
- e is a constant = 2.71

According to Murray, the constant β was chosen based on consultation with an advisory board to generate age patterns consistent with various “surveys and group exercises.”⁶¹ The constant C was chosen so that the introduction of unequal age weights would not change the global estimated burden of disease. Murray further claims that the relative arbitrariness of the value of β is

compensated for somewhat by the insensitivity of the global burden of disease to the value of β .⁶¹

“The important issue is not the exact form of an age-weighting function but the presence of non-uniform weights.” pg 61⁶¹

Barendregt *et al*^{7,91} contend that the age range emphasised by the age weighting system is not 9-54 as intended by Murray, but 0-27 years. This is because the Expected Years of Life Lost (EYLL) methodology used to calculate the DALY is an age-weighting system in itself that unduly compounds the formal age weighting. The impact of this bias will be tested in the subsequent example using DALYs.

4) Time preferences

Time preference is a component of the DALY because of the economic premise that individuals would rather benefit now than later. The discounting function of the DALY has an exponential form and the discount rate is assumed to be 3%^{39;74}.

Discounting function = $e^{-r(x-a)}$

Equation 6

Where:

- r is the discount rate, fixed at 0.03
- a = onset of disease
- x = age
- e is a constant = 2.71

Despite the use of discounting in cost-benefit and cost-effectiveness analysis, there is no consensus on the conceptual justification for discounting health benefits, or the appropriate discount rate for such benefits.^{12;32;36;44;45;63;64;67}

However, Murray ⁷⁴ puts forward the following arguments in favour of social time preferences (and the resultant discounting of health benefits):

1. Individuals may have a pure time preference due only to myopia.
2. If increased future consumption is accompanied by decreased marginal utility of consumption, then a marginal unit of consumption in the future will be less than in the present and the future unit should be accordingly discounted.
3. Time is related to uncertainty, therefore futures outcomes should be discounted to reflect the finite but positive risk that the individual will not exist at that time.
4. There is a possible time paradox which arises because one could argue that health benefits should be discounted at a lower rate than monetary costs and one is only indifferent to the time period when costs and benefits are discounted equally.
5. If health benefits are not discounted, we might erroneously conclude that all resources should be invested in disease eradication plans that have finite costs. This might crowd out interventions that do not result in disease eradication but which are efficacious none the less.

4.1) Discounting Life Expectancy and Intergenerational Equity

Anand and Hanson⁶ observe that due to the 3% per annum discount rate advocated by Murray^{97, 61} one life today would be worth 5 lives in 55 years time. This tendency to favour the current generation over future generations, weights treatments that offer to cure the current generation more heavily. This creates a bias against prevention programmes whose effects are only likely to be felt in the future. The methodological cause of this bias is the exponential nature of the discounting function.

The time preference argument behind discounting is defensible even in the health care environment where uncertainty is as likely to play a role as any other environment. Similarly, discounting can be defended on the grounds that any investment in health today is likely to yield a return in the future. However, while these arguments make intuitive sense, one could also see how the appreciation of current investments in health (or the decrements of future uncertainty) is unlikely to be exponential. It is for this reason – and from the resultant bias caused by exponential discounting - that writers in this field are beginning to advocate the presentation of economic evaluations both with and without discounting, as well as with sensitivity analysis' that test the sensitivity of findings to changes in the discount rate.⁵⁶

5) *Health can be summed across individuals*

The DALY calculation treats two people each losing 10 years of disability free life the same as one person losing 20 years. However, it is possible to weight duration non-linearly to give priority to fewer people suffering for long intervals.³⁹

6) *The complete DALY formula*

The number of DALYs lost due to disability at age x can be calculated using the following formula³⁹:

$$DALY(x) = (D)(Cxe^{-\beta x})(e^{-r(x-a)})$$

Equation 7

Where:

D = Disability weight ranging from 0 (perfect health) to 1 (death)^{§§§}

Equation 2 must now be integrated as follows, to calculate the number of DALYs lost from the onset of a disability (a) to the age of death (a+L) if the person lives to their maximum life expectancy.

$$DALY(x) = (D)(Cxe^{-\beta x})(e^{-r(x-a)})$$

$$\int_{x=a}^{L=a} DALY(x)dx = DC \int_{x=a}^{L=a} xe^{-\beta x - rx + ra} dx$$

and:

$$1) \quad u = x \rightarrow du = dx$$

$$2) \quad dv = e^{-x(\beta+r)} dx \rightarrow v = \frac{e^{-x(\beta+r)}}{-(\beta+r)}$$

therefore:

$$DC \int_{x=a}^{L=a} xe^{-\beta x - rx + ra} dx = DCe^{ra} \underbrace{\int_{x=a}^{L=a} xe^{-x(\beta+r)} dx}_{set=1}$$

Now Solving for I:

^{§§§} Recall that this means a DALY is a “bad” to be minimized.

$$\begin{aligned}
I &= uv - \int v du \\
I &= \frac{xe^{-x(\beta+r)}}{-(\beta+r)} - \int \frac{e^{-x(\beta+r)}}{-(\beta+r)} dx \Big|_{x=a}^{L+a} \\
I &= \frac{xe^{-x(\beta+r)}}{-(\beta+r)} + \frac{1}{\beta+r} \int e^{-x(\beta+r)} dx \Big|_{x=a}^{L+a} \\
I &= \frac{xe^{-x(\beta+r)}}{-(\beta+r)} + \left(\frac{1}{\beta+r} \right) \left(\frac{e^{-x(\beta+r)}}{-(\beta+r)} \right) \Big|_{x=a}^{L+a}
\end{aligned}$$

$$\begin{aligned}
I &= \frac{-e^{-x(\beta+r)}}{-(\beta+r)} \left(x + \frac{1}{\beta+r} \right) \Big|_{x=a}^{L+a} \\
I &= \frac{-e^{-(L+a)(\beta+r)}}{\beta+r} \left(L+a + \frac{1}{(\beta+r)} \right) + \frac{e^{-a(\beta+r)}}{\beta+r} \left(a + \frac{1}{\beta+r} \right) \\
I &= \frac{-e^{-a(\beta+r)} e^{-L(\beta+r)}}{(\beta+r)} \left(L+a + \frac{1}{(\beta+r)} \right) + \frac{e^{-a(\beta+r)}}{(\beta+r)} \left(a + \frac{1}{\beta+r} \right) \\
I &= -\frac{e^{-a(\beta+r)}}{(\beta+r)} + \left[e^{-L(\beta+r)} \left(L+a + \frac{1}{(\beta+r)} \right) - \left(a + \frac{1}{\beta+r} \right) \right]
\end{aligned}$$

$$I = -\frac{e^{-a(\beta+r)}}{(\beta+r)^2} [e^{-L(\beta+r)}(1 + (\beta+r)(L+a)) - (1 + (\beta+r)a)]$$

So,

$$\int_{x=a}^{L+a} DALY(x) dx = DCe^{ra} \int_{x=a}^{L+a} xe^{-x(\beta+r)} dx$$

$$= -\left[\frac{DCe^{-\beta a}}{(\beta+r)^2} \right] [e^{-(\beta+r)L}(1 + (\beta+r)(L+a)) - (1 + (\beta+r)a)]$$

This process of integration therefore yields the following formula.:^{39;91}

$$DALY = \left[\frac{(D)(Ce^{-\beta a})}{(\beta+r)^2} \right] [e^{-(\beta+r)L}(1 + (\beta+r)(L+a)) - (1 + (\beta+r)a)]$$

Equation 8

HIV Treatment in the Current Political Climate

HIV has become a highly politicised illness in South Africa, with the debate around mother-to-child transmission (MTCT) of the disease receiving considerable media attention. The latter half of 2001 however, appears to have seen a relaxing of government's anti-treatment stance following pressure from the Treatment Action Campaign and COSATU⁸⁹, and constitutional court challenge of the "no treatment" stance. According to the Minister of Health, between May 2001 and January 2002; 38168 women had been tested for HIV in public ante-natal clinics around the country. Of these, 9490 (24.9%) tested positive and 3734 (39.3%) received Nevirapine as part of the governments roll-out of a state package to reduce MTCT³⁰.

On the 17th of April 2002 government also agreed to offer ARV treatment to sexual assault survivors⁷². The MTCT programme and the provision of ARVs to sexual assault survivors are both measures that did not originally feature in the "HIV/AIDS and STI Strategic Plan for South Africa, 2000-2005." These measures have, however, been subsequently added to the plan and they illustrate that the government is prepared to be flexible to some extent about the best way to achieve the strategic objectives laid out in the document.

This flexibility is the motivation for the consideration of a highly active anti-retroviral treatment (HAART) package for AIDS sufferers in South Africa. Declining pharmaceutical prices and simpler treatment regimens may have put a basic treatment package within the bounds of cost effectiveness (or cost-utility in this case). The reader should note however, that despite the pragmatic motivation for this costing, it is merely presented here as an example of the DALY methodology at work. The interested reader is referred to the original paper by Boule; Kenyon; Skordis and Wood, which is pending publication in the

South African Medical Journal.¹⁰ The original paper provides the detailed costing of a HAART programme in South Africa. It accounts for mortality effects but not morbidity effects as the paper was intended as an advocacy tool. Morbidity effects are both more complex to calculate and to interpret and they can conceal more than they reveal, as the following example will show.

Back to Basics: A Review of ARVs in the HAART Context

In light of the media attention HIV/AIDS has received, and the well-publicised debate over the role of ARVs, it might be beneficial to review the generally accepted benefits and drawbacks of antiretroviral treatment in the HAART context before entering into the costing example:

- The HIV prevalence rate in a sexually active population will not exceed 20 – 40% at the peak of the epidemic.⁹⁶
- When approximately 2% of the sexually active population is infected, the spread of the disease becomes exponential.⁹⁶
- From the onset of exponential infection rates, it takes approximately ten years for prevalence rates to peak.⁹⁶
- Once the prevalence rates have peaked, the HIV is likely to become endemic to the affected society, with approximately 10% of the sexually active population being positive at any one time.⁹⁶
- Antiretroviral drugs (ARVs) prolong life but do not cure the disease.^{35,105}
- ARVs reduce the viral load of the patient thus reducing the extent to which they are infectious.^{24,73}
- Despite the improved outcomes, some problems such as failure and toxicity have been observed.³⁵

- There is a growing literature concerned with drug resistance following HAART, which can limit therapeutic options.^{31,35;105}

The pro-HAART argument revolves around the following main ideas:

- Access to medical care is a basic human right and ARVs have been show to be effective in prolonging the lives of AIDS sufferers.^{89;35;105}
- HAART can support the effectiveness of prevention campaigns.⁹²

The future treatment burden South Africa will inevitably face as a result of the HIV epidemic is beginning to emerge. Although the rate of new infections may have peaked, South Africa will be battling with increasing HIV morbidity for many years to come. While much of the public debate has thus far focused on marginal issues such as toxicity, there has been little constructive debate between opposing camps on the two key issues; cost and efficacy. Where the pro-camp have focused on efficacy and largely ignored issues around cost, the anti-campaign has dwelt on costs. Both are important issues and this exercise goes some way towards pinning down the true cost implications of a limited HAART program in South Africa.⁴²

Notwithstanding any future political decisions to mobilise extraordinary resources for HIV treatment, the following model provides the cost per DALY of a baseline national treatment programme. This example only makes use of the baseline scenario outlined by *et al.*¹⁰

Devising a unit cost of HAART treatment: An Overview of the Assumptions****

The following costing has been based on a service model which is currently emerging in the Western Cape services in which:

1. *“HIV/AIDS services are required to develop the relationships between patients and clinicians, to ensure continuity of care, and to provide a mechanism through which patients can be evaluated for potential enrolment onto an antiretroviral treatment programme. These services, whilst being part of a range of services offered at primary health care clinics, are not a replacement for the primary health care and often constitute an additional service, and should be costed as such.” pg 2¹⁰*
2. Consequent on meeting some form of eligibility criteria which do not contravene the patient’s constitutional rights to access, patients become eligible for HAART when they become AIDS symptomatic. This is generally coincides with the onset of Stage 4 of the HIV infection when HIV develops into AIDS.¹⁰² The baseline scenario used for this exercise allows for CD4 count testing which would further inform the initiation of treatment;¹⁰
3. It is assumed that patients would have their HAART program managed through a specialised HIV/AIDS service, although they would still attend conventional primary health care services for other needs.¹⁰

**** The model has been taken from a paper by Boulle *et al*¹⁰. The model was presented at the 4th International AIDS Conference in Barcelona in July 2002 and is pending publication in the South African Medical Journal.

Direct programme costs

Medicines

Medicine costs comprise the largest proportion of HAART program costs. This model uses the generic drug regimes and proportions detailed in *Table Three*. Furthermore, an annual drug price reduction is assumed which, in real terms, amounts to 10% per year for the first three years and 5% annually thereafter^{†††42}.

Table Three

	First Line Drugs for the Initial Year of Treatment	First Line Drugs From the Second Year of Treatment
	<i>Drug Name</i>	<i>Drug Name</i>
Single Tablets	D4T	D4T
	3TC	3TC
	NVP	NVP
		DDI
		EFZ
		INV/RIV
		AZT
Combination Tablets	Triamune	Triamune
		Combivir
Drug Cost per Year of Treatment	\$295	\$424

This drug regime has been designed to allow at least two independent regimens, whilst still sufficiently limiting costs. It is assumed that all rationally selected HAART regimens have equivalent treatment outcomes.⁸ Where a combination tablet could substitute for individual medicines, the price of the combination tablet was included due to possible adherence advantages.¹⁰

††† Prices have dropped by a greater margin than this in the last year in South Africa. Pressure by the National Department of Health, various AID agencies and NGO's such as the Treatment Action Campaign is likely to result in a continuation of these price declines.

Diagnostics

Based on the WHO recommendations⁹¹ the baseline costing model incorporates all tests (including a biannual CD4 count and a CD4 count prior to enrolment) with the exception of viral load tests.¹⁰

Visit costs for the antiretroviral treatment programme

The lack of published cost data for public-sector ARV treatment programmes makes the cost of HIV treatment for the health sector particularly difficult to estimate. However, given the reasonable assumption that clinical costs may be higher for HAART than for standard primary care consultations, the costing model applies a cost factor of 1.5 to the average primary care consultation costs in the Western Cape metropolitan area, where there is a doctor-driven primary health care service.^{###10}

A visit schedule based on the current Médecins Sans Frontières (MSF) treatment protocol in the Western Cape^{52,53,91} is used to estimate the additional visits required to sustain HAART. The baseline visitation scenario in Boulle *et al.*¹⁰ assumes that the majority of visits have a lower cost structure than a conventional doctor's consultation as the visit is principally to ensure adherence and dispense medicines.^{§§§§} The model further assumes that existing primary health care needs are provided for by the routine health services and are not costed as part of the intervention.

It is plausible to assume that a proportion of Stage 3 HIV sufferers will make use of HIV/AIDS services i.e. before becoming AIDS-symptomatic. The baseline model sets this proportion equal to the proportion of patients accessing treatment when becoming AIDS symptomatic. This results in a higher number

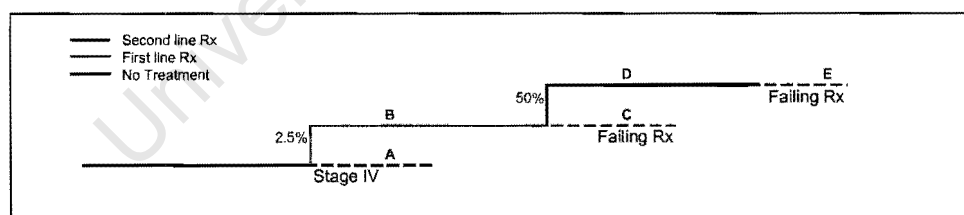
The full model includes a sensitivity analysis on this parameter

§§§§ Boulle *et al.*¹⁰ also include another visitation scenario in which a greater proportion of the visits takes place with a doctor and therefore has a higher cost structure. More detailed sensitivity analysis is also presented in the paper.

of visits to the HIV/AIDS services by people who are not on antiretrovirals (ARVs) in the initial years, reducing as a proportion of all visits over time as the numbers of people on ARVs accumulate. The model assumes 3 visits per year on average for those not on ARVs but attending the HIV/AIDS services as a prelude to possible enrolment into an ARV treatment programme. These visits are costed at the existing PHC consultation costs.¹⁰

The ASSA2000 AIDS and Demographic model² was used to anticipate the numbers of HIV-infected people over the next ten years. This distribution was further segmented by WHO clinical stage to identify the likely numbers of Stage 3 and Stage 4 sufferers. A proportion of those who would otherwise become AIDS symptomatic (Stage 4) were assumed to move onto antiretroviral treatment. Probability distributions based on the median survival rate were used to determine the probability each year of changing treatment (only one regimen change is provided for in the baseline scenario), failing treatment, and dying based on the duration of treatment (see *Figure One* below).

Figure One ****



LY gained per non-failing year on first line treatment: $(B+C-A) / B = 0.85$

Source: Slides from Barcelona Presentation¹⁰

**** In this model, Rx refers to a treatment regime.

Placing programme costs within the DALY framework

The sub-populations determining programme costs for the complete model are those who are on first-line treatment, those on second-line treatment, and those failing treatment. However, for the purposes of this example, we are only interested in the proportion of people who are on first line treatment and the proportion of those people who fail treatment (B and C in Figure One).

This model yields a unit cost of approximately R5,600 (\$562.81^{††††}) per person, per life year saved. Putting this unit cost into the DALY calculation framework (which is repeated below) yields the results summarised in *Table Four*.

The DALY formula:

$$DALY = \left[\frac{(D)(Ce^{-\beta a})}{(\beta + r)^2} \right] \left[e^{-(\beta+r)L} (1 + (B+r)(L+a)) - (1 + (r + \beta)a) \right]$$

†††† This is calculated at an exchange rate of R9.95 to \$1

Table Four

Standard DALY calculation							
THE CASE WITH HAART				THE CASE WITHOUT HAART			
<i>First calculating the number of DALYs lost to disability while in Stage Four of AIDS sickness</i>							
Variable name	Variable Description	Source of Estimate	Value	Variable name	Variable Description	Source of Estimate	Value
C	Constant	³⁹	0.16243	C	Constant	³⁹	0.16243
D	Disability weight ^{###}	^{14,35,39}	0.2215	D	Disability weight	^{14,35,39}	0.505
r	Discount weight	^{39,61}	0.03	r	Discount weight	^{39,61}	0.03
β	Constant	³⁹	0.04	β	Constant	³⁹	0.04
a	Age when disability starts	¹⁴ (average age at HIV infection)*	25	a	Age when disability starts	¹⁴ (derived)	25
L	Years lived with the disability	^{42,91}	5	L	Years lived with the disability	^{42,91}	1
e	Constant	³⁹	2.71	e	Constant	³⁹	2.71
DALYs lost =			1.53	DALYs lost =			0.74
<i>Now calculating the number of years lost to premature mortality:</i>							
C	Constant	³⁹	0.16243	C	Constant	³⁹	0.16243
D	Disability weight	^{14,39}	1	D	Disability weight	^{14,39}	1
r	Discount weight	^{39,61}	0.03	r	Discount weight	^{39,61}	0.03
β	Constant	³⁹	0.04	β	Constant	³⁹	0.04
a	Age when the person dies	^{42,91}	30	a	Age when the person dies	^{42,91}	26
L	Years of life lost due to premature death	⁶¹	53.27	L	Years of life lost due to premature death	⁶¹	57.17
e	Constant	³⁹	2.71	e	Constant	³⁹	2.71
DALYs lost =			29.40	DALYs lost =			31.66

Summary Figures from Table Four:

Total DALYs lost due to HIV when on treatment	30.93	Total DALYs lost due to HIV without treatment	32.40
--	--------------	--	--------------

<i>Total DALYs averted by treatment:</i>	1.4800
<i>Cost per DALY (R5,600 per year for 5 years divided by the total DALYs averted):</i>	R18,970.57
<i>Cost per DALY in dollars**:</i>	\$1,906.69

This disability weight is calculated on the assumption that a person is symptomatic when they present for treatment (thus having a disability weight equal to that of patients not on treatment (0.505)). Patients may not seek treatment immediately and treatment is not immediately effective so it is assumed that this weight is maintained for 6 months. The patient is then assumed to have a disability weight close to zero (0.1) for 3.5 years before they start to fail on the medication and are once again assigned a disability weight of 0.505 for the final year of life.

Using these figures as a basis, one can manipulate the assumptions of the DALY model to appreciate the sensitivity of the calculation to changes in the assumptions. The discount rate, the life expectancy assumptions and the age weighting function are the three parameters which will now be varied in turn.

Changing the discount rate...

In the first instance the discount rate has been reduced to zero.^{6:56,91} Reducing the discount rate to zero increases the number of DALYs averted by treatment, this decreases the program's cost per DALY and increases the implied cost effectiveness.

Reducing the discount rate to zero:

Total DALYs lost due to HIV when on treatment	53.27	Total DALYs lost due to HIV without treatment	58.25
--	-------	--	-------

<i>Total DALYs averted by treatment:</i>	4.9795
<i>Cost per DALY (R5,600 per year for 5 years divided by the total DALYs averted):</i>	R5,623.03
<i>Cost per DALY in dollars**:</i>	\$565.13

In the second instance the discount rate has been increased to 10%.⁶¹ As one would expect, this manipulation has the effect of lowering the number of DALYs averted by the project and hence dramatically raising the cost per DALY of the program.

Increasing the discount rate to 10%:

Total DALYs lost due to HIV when on treatment	14.30	Total DALYs lost due to HIV without treatment	14.33
--	-------	--	-------

<i>Total DALYs averted by treatment:</i>	0.0353
<i>Cost per DALY (R5,600 per year for 5 years divided by the total DALYs averted):</i>	R 792,255.82
<i>Cost per DALY in dollars**:</i>	\$79,623.70

Changing the life expectancy assumptions...

As this costing applies only to the South African context, it is interesting to consider the impact on the DALY of using the South African life expectancy rather than the higher, 'ideal' life-expectancy assumed by the standard DALY calculation. According to the World Bank, South Africa's life expectancy for 2000 was 47.8 years.¹⁰¹ Using a simple calculation, the years of life lost are calculated as 47.8 less the age at death. The effect of this was somewhat surprising, instead of making the program less cost-effective, it makes the program more cost-effective. This occurs because the difference between the numbers of DALYs with treatment and the number without treatment increases. Of course, this may simply be the case with this particular intervention. It is not suggested that this will always be the case, however this finding makes the point that differences in life expectancy assumptions have implications for CUA findings.

Total DALYs lost due to HIV when on treatment	20.02		Total DALYs lost due to HIV without treatment	22.73
--	-------	--	--	-------

<i>Total DALYs averted by treatment:</i>	2.7152
<i>Cost per DALY (R5,600 per year for 5 years divided by the total DALYs averted):</i>	R 10,312.44
<i>Cost per DALY in dollars**:</i>	\$1,036.43

Changing the age weighting function...

$$\text{Age weighting function} = Cxe^{-\beta x}$$

As discussed earlier, there was some arbitrariness to the assignment of values for β and C. According to Murray, the value of β could in fact lie anywhere between 0.03 and 0.05 – admittedly narrow bounds.⁶¹ Setting β equal to the lower bound has the effect of substantially increasing the burden of disease (the number of DALYs lost with and without treatment). It also reduces the “effect” of the intervention which increases the cost per DALY averted.

Total DALYs lost due to HIV when on treatment	48.95	Total DALYs lost due to HIV without treatment	49.97
--	-------	--	-------

<i>Total DALYs averted by treatment:</i>	1.0143
<i>Cost per DALY (R5,600 per year for 5 years divided by the total DALYs averted):</i>	R 27,604.95
<i>Cost per DALY in dollars**:</i>	\$2,774.37

DALYs are designed to favour the economically active (and those who are age-eligible to be economically active).⁶¹ This implies a human capital approach to human worth⁵⁶ which should be treated with caution. As the above example shows, this human capital approach can lead to an underestimation of the burden of disease in a community. Particular care should be taken in a country like South Africa which has been hit hard by the HIV pandemic. AIDS changes the age structure of the population resulting in an increase in child-headed households^{19,91} and pensioners supporting many grandchildren.^{19,91} People in countries like South Africa no longer have an equal chance of living to every age and one could argue that the DALY is biasing against the very age groups who are likely to hold the fabric of society together in the face of spiralling adult mortality amongst the working-aged (and sexually active) population.

Furthermore, if one can justify structuring the DALY to favour the working aged, then surely that same argument can extend to favouring the providers of essential services such as doctors and nurses, and the employed over the unemployed. In short, the age weighting implies social preferences that may not in fact be true.

A FINAL WORD ON DALYS

Parker and Fox-Rushby believe that it is inappropriate to assume that dimensions of health are identical across differing cultural environments.⁶⁶ They have three criticisms of generic methods such as the DALY:

1. Firstly, most generic instruments were developed in wealthier countries such as the USA and Western Europe and as such they emphasize the inextricable link between health, well-being, individuality and autonomy. Parker and Fox-Rushby point to a large body of anthropological literature which suggests that in non-Western cultures, the different aspects of an individual's health may bear little relation to the way that individuals perceive their health and well being within a community.
2. Secondly, all of these generic instruments are inevitably cultural instruments, yet the associated literature does not pause to consider what culture is.
3. Finally, generic methods are constructed with scant attention to qualitative methods of data collection such as participant observation supplemented by unstructured and semi-structured interviews over time. These methods could add considerable value to the understanding of quality of life.

Parker and Fox-Rushby return us to the trade off mentioned earlier between accuracy of measurement within an homogenous context and comparability of measurements across heterogeneous contexts. They question the importance of international comparisons and urge regional decision makers to consider the cost-effectiveness of developing tailored instruments for a unique population.

Garratt *et al*²⁸ lend weight to the argument against generic measures of burden of disease. Their study concludes that in some specialities, there are numerous measures of quality of life and little standardisation. Of those patient-assessed measures of quality of life that were classifiable under the methodology of their study, 46% were disease or population specific, 18% were dimension specific, 10% were utility and 1% were individualised measures. Only 22% of patient assessed quality of life measures could be classified as generic.

At best the DALY might assist in optimising allocative efficiency within the health budget but it cannot assist in optimising broader allocative efficiencies in society. Cost utility analysis, as a sub-discipline of CEA, falls victim to the same key criticism: While CUA can assist in the ranking of health interventions, CUA cannot advise a cut-off point or optimal program size for each intervention. Nor can CUA say anything about the positive impact that one intervention can have on another.⁵⁴

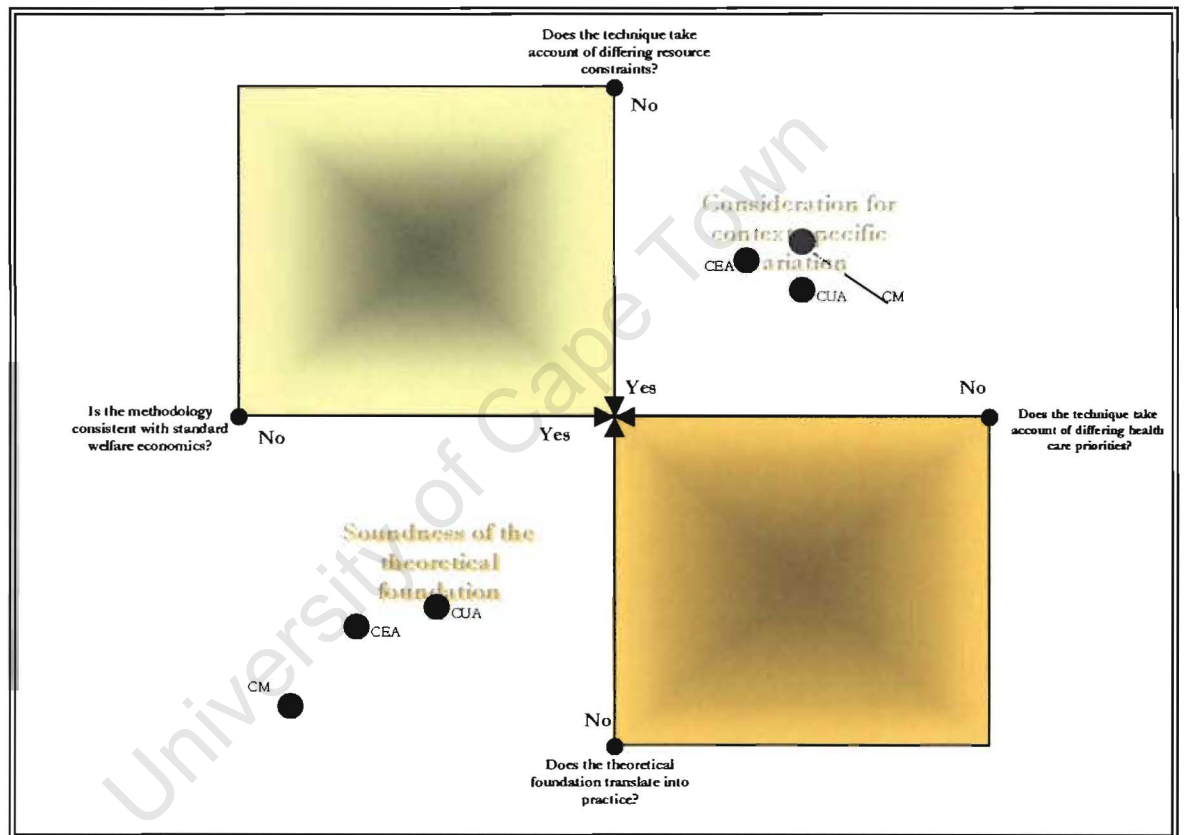
Furthermore, CUA and particularly the DALY can be guilty of obscuring more than it reveals. As the example above shows, changes in the discount rate, the life expectancy assumptions and the age weighting function can dramatically impact on the results and hence the practical implications of the model. Normative decision rules are being concealed within an apparently complex, scientific procedure. Those normative decision rules could therefore be manipulated to suit the ends of special interest groups – and that manipulation is given certain credibility within the DALY framework.

By adding considerations of morbidity to those of mortality, the DALY has a valuable contribution to make; however, steps should be taken to ensure consistency in the application of assumptions and transparency regarding the implications of those assumptions. In countries where the institutions for health

delivery have fewer checks and balances, this consistency and transparency is particularly crucial.

Attempting to place CUA within the formal assessment matrix then:

Figure Two: Positioning CUA within the assessment matrix



*Chapter Six: Cost Benefit Analysis – The Human
Capital Approach versus Willingness to Pay*

INTRODUCTION

Cost benefit analysis attempts to monetise the costs and benefits of a health intervention.^{40;91;95} The two principal methods for measuring health care benefits in a cost benefit analysis are the willingness to pay approach and the human capital approach.⁹¹ As the following discussion will show, both of these methods are particularly challenging to apply in developing countries. However, an emerging methodology known as the “net-cost methodology”, may offer some direction for a new approach to measuring benefit⁸³. An approach more suited to the developing country context.^{18;20;33;40;76;79;80;84;88;95}

THE HUMAN CAPITAL APPROACH

The human capital approach measures the benefit of an intervention in terms of an individual’s future stream of income. The value of one life year is equated to an individual’s annual salary and as such the value of a life saved or extended, is the sum of that income over the extended lifespan of the patient.⁹¹ Even this rather succinct explanation of the human capital approach raises an immediate problem: how does one value benefit for those who do not earn a salary?^{20;40;82;95}

The human capital methodology raises ethical concerns in countries with high unemployment rates. Similar concerns are raised when valuing interventions for women or children who are less likely to be economically active. A human life is valued only in terms of the money earned over its duration and this would serve as the objective function in an optimisation problem. Obviously this raises considerable moral and ethical objections, for example, is an honest low wage earner a less valuable individual than a high earning white collar criminal? This

example may seem extreme but it makes the point that the value of individuals is not necessarily captured by the sum of their pay cheques.

“...the human capital approach has several limitations. There is a lack of theoretical foundation, the intrinsic value and quality of life are ignored and the method discriminates against people outside the labour force.” page 2^{40,91}

Ethical dilemma's aside, the human capital approach raises a number of technical difficulties such as how to treat retired persons and women who work in the home.^{82,91} Even among the employed the choice of an appropriate wage rate can cause a problem in the absence of comprehensive earnings data – a problem not uncommon in many developing countries. How does one decide on an appropriate minimum or standard wage for various forms of employment?^{82,91}

A study conducted in Israel^{29,91} offers one possible solution to these difficulties. Instead of looking at the salaries of the particular beneficiaries of a health intervention, these authors simply use annual GNI per capita. This immediately levels the playing field between the employed and the unemployed, the economically active and the inactive. All members of that society are valued equally. This circumvents many of the ethical problems implicit in the human capital approach but does not address the concern that a health intervention is not in place solely to maximise GNI. This solution would also not aid an institution such as the World Health Organisation who may need to make cross-country comparisons as it would result in expenditure being biased towards wealthier countries – a phenomenon which makes little sense in an aid context. Thus, while this solution partly addresses intra-country comparisons, it does not solve the difficulty of inter-country comparisons.

On the positive side, the human capital approach allows one to monetise directly the benefit of an intervention in a simple manner if the data is available. It does

not require lengthy questionnaires and there is little risk, if any, of respondent bias or over claim.^{82,91} Input data is more readily available in published form than that for the willingness to pay approach and as such requires less time and expense to compile.^{82,91} However, Johanneson and Jonsson⁴⁰ point out that despite this appeal, *“If the human capital approach was routinely used as a basis for allocating health care resources, important misallocation of resources would result,”* (page 3). They see the only advantage of the human capital methodology is the contribution it has made to the calculation of direct and indirect medical costs. This was an important precursor to the development of cost-effectiveness analysis.^{40,91}

WILLINGNESS TO PAY

Very simply, the willingness to pay approach estimates the amount that an individual is prepared to pay to receive a particular health increment (i.e. not to receive perfect health or death).^{57,79,80,88,91} Common methods for measuring willingness to pay include the standard gamble, time trade off methods and standard analogue scale,^{§§§§§,13,75,82,88,91} which fall under the ambit of contingent valuation (CV).^{57,91} Conjoint analysis can also assist in the establishment of willingness to pay, as can the observation of risk avoidance practises such as the amount paid on life insurance or for cars with airbags and other safety devices. Each of these methods try in different ways to identify the amount of value placed on a given health increment (or the avoidance of a health decrement).^{82,91}

The willingness to pay approach is the most consistent with social welfare theory.^{20,40,57,91} By establishing how much an individual is willing to pay to gain a health increment (or avoid a health decrement) willingness to pay attempts to

§§§§§ A more detailed explanation of these techniques falls outside the constraints of this paper. For explanations of these and other techniques, the reader is referred to the texts cited.

price the intervention at the point where the individual's marginal utility from the intervention is equal to the marginal cost of the intervention to the same consumer thus ensuring allocative efficiency.

The willingness to pay methodology attempts to derive a convex utility function such as that commonly shown in Edgeworth box analysis and in so doing arrives at a Pareto efficient outcome with an accurate shadow price, undistorted by market failure and constrained only by the service provider's capacity to delivery.^{20;46;91} This can be done using the methods of compensating variation or equivalent variation.^{40;43;91}

"If we are investigating willingness to pay for a specific treatment, willingness to pay is equal to the Hicksian compensating variation²¹ if the consumer is not initially undergoing treatment, and is equal to the Hicksian equivalent variation if the consumer is undergoing treatment." pg 8^{40;91}

In practice, a number of complications arise. For example, how is someone who currently does not suffer from asthma or HIV supposed to value asthma or HIV treatment? How is someone who has lived all their life under a social health care regime (provided and paid for by the government) and has never paid for health care supposed to understand realistic prices for a range of interventions? These questions are particularly pertinent to the developing country context and reflect the oft cited difficulty of measuring health changes in monetary terms.^{40;58;59;61;79;80;84;91;95}

This difficulty relates directly to a further criticism of the willingness to pay approach: the burden of obtaining the necessary data to compute willingness to pay measurements^{20;40;57;91}. Willingness to pay for health interventions is obtained from survey data. Questions can be asked in a number of ways – either by simply

²¹ The reader is referred to Kreps^{43;91} for a detailed explanation of the theory of compensating and equivalent variation and Hicksian functions.

asking how much a person would be willing to pay for a certain risk reduction^{40,95} or by asking the respondent to select between pairs of successive choices until the respondent is indifferent to the trade-off presented.^{20,91,95}

The second alternative raises a particular challenge in the developing country context as the researcher needs to generate a viable set of alternatives in environments where few treatments are likely to exist for any one ailment. Regardless of how the question is asked, a number of other problems are likely to arise. In South Africa for example, the existence of eleven official languages poses an immediate problem, which language should the instrument be asked in? And does the question stand up to the translation? A further problem is the nature of interviewing conditions in many poorer or rural parts of the country where crowded living and long distances between respondents complicate the process and increase the expense. Combine this environment with the usual respondent fatigue, add too the fact that the respondent is unlikely to understand how the information will be used, combine these shortcomings with the fact that the true risk of the tradeoffs presented may not be understood by the respondent – and one could rightly question whether the resultant data is likely to reflect the respondent's true utility function.

Another frequent criticism of willingness to pay is its unavoidable reflection of ability to pay^{79,80,91} – \$10 to a wealthy person is very different to \$10 to a person living below the poverty line of \$1 dollar a day. A wealthy person may be prepared to pay a substantial sum to reduce their risk of heart disease while a poorer counterpart may be more concerned with the immediate problem of feeding a family. The risk of heart disease at some future date is highly unlikely to be given considerable share of mind, or share of pocket, when the threat of starvation is imminent.

In a society with a history of free (or reduced cost) health service provision, over claim is a particularly common problem with willingness to pay.^{61,91} Respondents do not believe they will ever have to pay for the intervention they are valuing but they gain some positive utility from knowing that the service or intervention is available should they ever need it.^{57,79,80,88,91} As such, they over-value the service to increase the chances of its provision. Over-claim may also result in cases where it may be considered socially desirable to be prepared to pay all you have. For example, when being interviewed in a crowded living room about how much you would be willing to pay for an intervention to prevent measles in your children you may feel as though you should be prepared to pay anything to prevent the illness – particularly if your mother in law is sitting next to you!

A further problem with willingness to pay is the lack of an acceptable benchmark. How much is too much? How much is too little?^{57,91} Respondents struggle to answer these questions and analysts frequently find themselves with the same dilemma.

On the positive side, willingness to pay correctly conducted enables one to understand and even to rank individual and social preferences in a way that the revealed preferences methodology does not.^{57,88,91} For example, if an individual is willing to pay more to avoid waiting longer at a clinic than to receive more modern forms of treatment, this ordinal ranking, quite aside from the cardinal differential, allows one to allocate resources more consistently with individual and hence social preferences. Morrison and Gyldmark^{57,91} cite Gafni's^{24,91} three criteria for successful contingent variation that adheres to the requirements of social welfare theory:

1. The respondent must be informed of the probability that they will ever need the service being valued. Willingness to pay questions must then be phrased so as to ask for the maximum amount that the person would pay

as an insurance premium to ensure that the service is available should they ever need it.

2. A separate imperative is that the patient understands the probability of successful treatment if a treatment is ever undertaken.
3. Finally, as willingness to pay should represent the willingness of a population, care should be taken to ensure the survey sample is representative of the treatment population.

Morrison and Gyldmark use a number case studies to argue that, thus far, *“these criteria have not generally been fulfilled in health-related CV studies and, even if they had been, the criteria themselves are not sufficient to obtain valid and useful results from CV studies.”* page 237^{57,91}

So, while willingness to pay may offer some advantages if correctly conducted, the informational and procedural requirements for that correct implementation may fall outside of the resource constraints of the average developing country.

ALTERNATIVE APPROACHS TO MEASURING BENEFIT

It appears that little effort has been invested in finding alternative ways of measuring benefit and the willingness to pay approach remains popular in the literature. Murray and Salomon⁶² investigated the use of mathematical modelling to evaluate global TB control strategies but their approach cannot yet be generalised to other cases. Another approach which holds promise is that used by Natrass in a polemic piece on mother to child transmission of HIV/AIDS in South Africa. Skordis and Natrass⁸³ further developed what has become known as the “net-cost methodology” and effectively applied it to a formal economic evaluation of mother to child transmission (MTCT) of HIV in

South Africa and the viability of an MTCT intervention such as AZT or Nevirapine.

The Net-Cost Methodology

In brief, the net-cost methodology weighs the costs of a health care intervention to the provider, against the costs to the provider of “not intervening”. The mother to child transmission case provided the perfect testing ground with the benefits to the public health care budget clearly outweighing the costs to the budget (or, put differently, with the cost of not intervening being higher than the costs of a basic MTCT intervention).⁸³ However, the MTCT case was one of prevention rather than palliative care. Palliative care can be a very different health context to prevention but Boulle *et al*⁸⁰ have illustrated that the net-cost methodology can make a meaningful contribution to palliative care costing.

To illustrate, we pick up on the example used to calculate sample DALYs in Chapter Four. The costing, as before, is for a baseline HAART program in South Africa and the results of the study have been summarised in *Table Three* below. These findings show that it is cost-saving to the health budget to offer HAART.

Table Three

Total Hospital, PHC and TB costs in 2007 with no ARV intervention	R15,988,193,906
ARV program costs in 2007	R140,302,366

Total Hospital, PHC, ARV and TB costs in 2007 WITH ARV intervention	R15,970,134,904
Accumulated difference over 5 years between nil and baseline scenarios	R-59,150,482
Proportion of ARV costs potentially offset at given time-point	119%

Source: Slides from Barcelona Presentation¹⁰

Total Hospital, PHC and TB costs in 2007 with no ARV intervention have been calculated using the following assumptions (see Table Four):

Table Four

	<i>Units</i>	<i>Stage I&II</i>	<i>Stage III</i>	<i>Stage IV</i>	<i>ARV FL</i>	<i>ARV SL</i>	<i>Failing</i>	<i>Unit cost Min</i>	<i>Max</i>
Tuberculosis	<i>Annual Risk</i>	5%	17%	34%	2%	2%	34%	\$420	
Hospitalisation	<i>Inpatient days</i>		5.8	18.8	1.5	2.0	16.5	\$45	
PHC consultations	<i>Extra consults</i>		8	13	2	2	13	\$7	
ARV medicine costs	<i>Annual cost</i>				Full	Full	Half	\$421	\$1,023
ARV laboratory costs	<i>Annual cost</i>				Yes	Yes	Yes	\$37	\$156
ARV visit costs	<i>Visits</i>				12	12	3	\$82	\$88

Source: Slides from Barcelona Presentation¹⁰

As with any model it is necessary to make assumptions and simplifications. One of the unfortunate – but necessary - simplifications of this model is that it only takes into account the costs of HAART to the public health system. However, it

could well be argued that HAART reduces the need for spending in other areas.

For instance:

- Amongst the employed, HAART may help to improve (or at least maintain) the productivity of HIV positive workers and prolong their working lives thus reducing staff turnover and new staff training costs. This might comprise a substantial saving to both the private and public sectors.
- There is also a growing concern in society about a large cohort of AIDS orphans¹⁹ who may themselves be either HIV positive or negative. Keeping these childrens' parents alive might help to keep this cohort of children off the welfare system and out of state orphan care facilities. This may comprise a significant saving to the Welfare Budget.
- Obviously, HAART does not cure AIDS but significantly delays the onset of debilitating illness and death. One could argue that this delay will have a discounting effect on the subsequent treatment cost once the patient fails on the treatment. This benefit is quite aside from the fact that five years is a long time in medical-development terms. Who knows what new drugs, vaccines or cures may be developed in that period? Buying time is buying hope.

The net cost methodology, as it was employed Boulle *et al*²⁰, fails to quantify these potentially substantive elements. The primary reasons for this are twofold:

1. Firstly, the data requirements are substantial and, in some cases, costly to obtain.
2. Secondly, capturing some of these benefits would require the quantifying social norms and preferences. This would effectively

revert us to the arguments against the assumptions implicit in the DALY calculation, that social norms differ too greatly from community to community to be easily quantified.

The net cost methodology holds considerable potential but it needs further thought and investigation. Measuring the benefit of health interventions is arguably the area of the economic evaluation of health care which needs to most work.

CONCLUSION

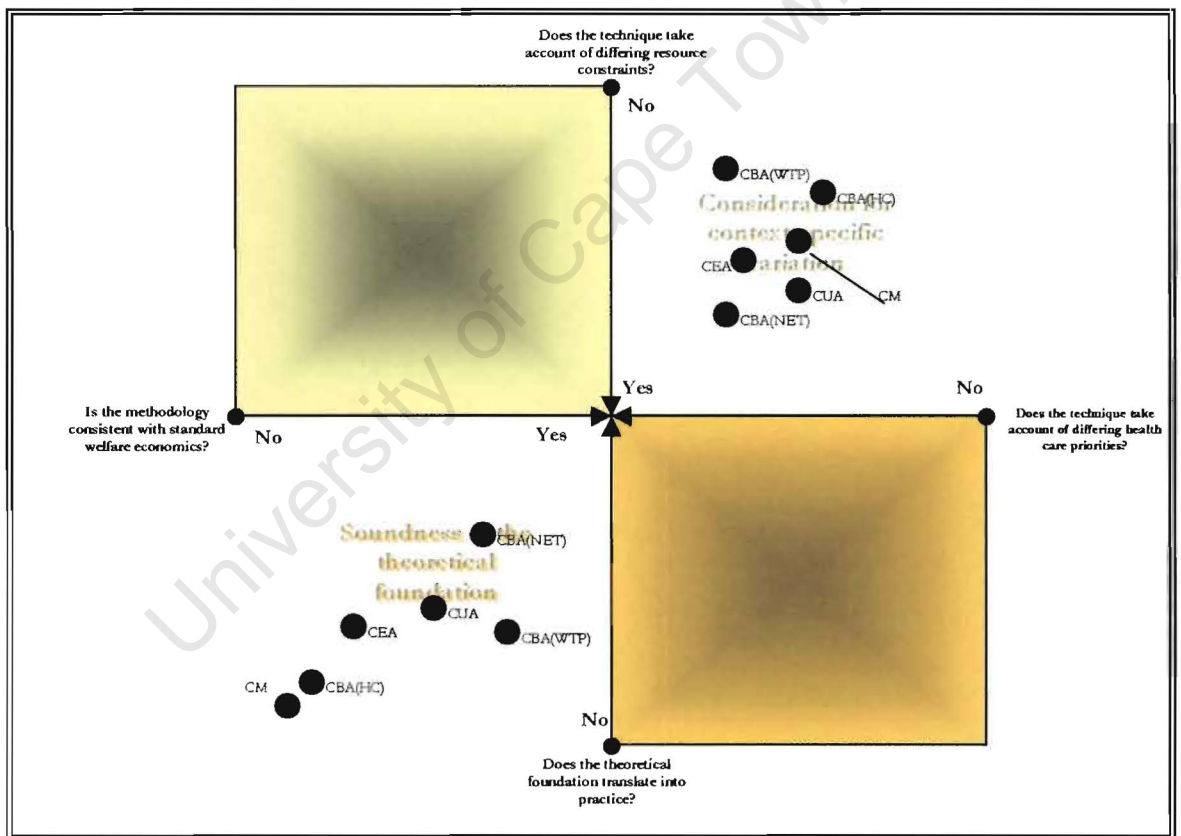
To summarise and attempt to position CBA on the assessment matrix.:

- The human capital approach has very poor theoretic foundations.⁴⁰ It is also particularly unsuited to the developing country context because of its bias against the unemployed and the measurement of human value in terms of streams of income.
- The willingness to pay approach is consistent with social welfare theory.^{20,40,57} However, this firm foundation does not translate well into practice in the developing country context because of the relationship between willingness to pay and ability to pay, and the considerable data requirements for measuring willingness to pay. The same factors that prevent the theory translating into practice, make willingness to pay unsuited to the developing country context.
- The net-cost methodology is best able to account for the allocative implications of health spending both within the health budget and between health and other budgets. The net cost methodology does take into account a country's ability to pay for a programme however, it

suffers from the same problem as willingness to pay in that the data requirements are considerable (and are often prohibitive). Only these data constraints prevent the net-cost methodology from being very well suited for the developing country context. Skordis and Nattrass⁸³ have clearly shown how the methodology can assist in setting health priorities.

Placing the various forms of CBA within the formal assessment matrix then:

Figure One: Positioning CBA within the assessment matrix



Chapter Seven: Conclusion

Cost minimisation, cost effectiveness, cost utility and cost benefit analysis are all able to contribute to the design and implementation of health policy. The key caution of this paper is not to overstate that contribution. Developing countries face particular resource constraints and different health priorities to their wealthier counterparts. Resource constraints are often further exacerbated by information asymmetries between voter and politician as well as less than perfect markets. The fact that many of the techniques discussed above consider one health problem in isolation, without considering the broader implications of the expenditure can be particularly problematic in the developing country context. As mentioned before, if the opportunity cost of the investment has not been accounted for, concrete policy recommendations would be ill advised in any setting. Many of the shortcomings attributed to the methodologies discussed above might be overcome if the economic evaluation were conducted in a socio-political environment in which issues such as budget allocations could be considered without the undue influence of special interest groups and influential institutions. Again, as mentioned before, this is seldom the case in the developing country context where new democracies are cautiously finding their way and single party states or military dictatorships are not uncommon.

Outside of perfect competition in all markets, the Pareto conditions for a social welfare optimum cannot apply and the policy maker is left to satisfice in the world of the second best. Satisficing usually necessitates the application of normative decision rules and this is where the economic evaluation of health care has value. By quantifying the economic implications of a health intervention in some appropriate way (even within the narrow context of the health budget as is the case with cost effectiveness analysis and cost utility analysis), economic evaluation can place this normative decision-making in a more transparent light.

Inappropriate evaluation techniques (or even the inconsistent reporting of appropriate techniques) can however, have the opposite effect. In cost utility analysis, changes in the discount rate, age weighting assumptions, disability weighting assumptions and other variables can all effect the ultimate findings and hence, the recommended policy response. Explicit sensitivity analysis and more consistent treatment of assumptions will help to improve the value of economic evaluation in all contexts.

These changes are, however, largely cosmetic. They do not address the key trade-off mentioned at various points throughout this paper: namely the need for indicators that can be compared across differing contexts versus the need for a measure which accurately measures a problem within a particular context.^{65,91} Composite BoD indicators such as the DALY are particularly guilty of favouring generic measurement over accurate, context-specific measurement. This bias has implications for recommended policy responses.

Finally, care should be taken when applying the human capital and willingness to pay methodologies in the developing country setting. These methodologies can result in inequitable biases which, particularly in the case of South Africa, might be considered a contravention of human rights. The quote below, extracted from personal communication with a human rights lawyer, sets the tone for health care delivery, and the delivery of all public services, in South Africa:

“Section 9(1) of the Constitution specifically recognises that everyone is equal before the law and has the right to equal protection and benefit of the law. Section 9(3) proceeds to prohibit unfair discrimination, whether direct or indirect on a host of grounds. The prohibition of unfair discrimination refers to both overt forms of discrimination as well as criteria which may be neutral on its face but, ultimately results in certain groups being disproportionately affected.” From personal communication with Karrisha Pillay, human rights lawyer.⁶⁹

The reader is reminded that the spatial differences between methodologies on the comparative matrix are the result of a qualitative assessment only which is, of necessity, a conceptual and hence a subjective framework. The specific coordinates on the matrix aside however, the net-cost methodology appears to offer an alternative way of measuring health benefits which has not yet been fully exploited. It is possible that with further development, this may be the most theoretically sound and context-sensitive methodology of those assessed above. That said however, it is also possible, that the need to make assumptions about costs in order to reduce all variables to an equivalent monetized value could render the net-cost methodology vulnerable to the same criticisms levelled against other methodologies.

This paper thus serves to highlight two areas for future research. The first is in the development and testing of the net-cost methodology. The second is to meld the conceptual comparative framework into a more rigorous and objective form. Both of these areas will, however, require the generation or collection of data not currently available and as always, this may be a simpler proposition to make in theory than in practise but the challenge is laid out none the less.

Bibliography

1. Abedian I, Strachan B, Ajam T. *Transformation in action: Budgeting for health service delivery*. Cape Town: University of Cape Town Press, (1998).
2. Actuarial Society of South Africa. ASSA2000lite. <http://www.assa.org.za> (Downloaded from the internet on 20/07/02)
3. Anand S, Hanson K. Disability-adjusted life years: a critical review. *Journal of Health Economics* (1997); **16**:685-702.
4. Anand S, Hanson K. DALYs: Efficiency Versus Equity. *World Development* (1998); **26**.
5. Anand S, Hanson K. Disability-Adjusted Life Years: A Critical Review. *Journal of Health Economics* (1997); **16**:685-702.
6. Anand S, Hanson K. DALYs: Efficiency Versus Equity. *World Development* (1998); **26**:307-10.
7. Barendregt JJ, Bonneux L, Van der Mass PJ. DALYs: The age weights on balance. *World Health Report* (1996).
8. Bartlett JA, DeMasi R, Quinn J, Moxham C, Rousseau F. Overview of the effectiveness of triple combination therapy in antiretroviral-naïve HIV-1 infected adults. *AIDS* (2001); **15**:1369-77.
9. Bayoumi AM, Redelmeier DA. Preventing mycobacterium avium complex in patients who are using protease inhibitors: a cost effectiveness analysis. *AIDS* (August 1998); **12**:1503-12.
10. Boule A, Kenyon C, Skordis J, Wood R. Rationing HAART Part I: An exploration of the costs of a limited public sector antiretroviral treatment programme in South Africa. Publication pending, (2002).
11. Brown CV, Jackson P.M. Budgets, Bureaucrats and Efficiency. *Public Sector Economics* (1994); 193-222.
12. Cairns JA. Valuing future benefits. *Health Economics* (1994); **3**:221-9.
13. Cook J, Richardson J, and Street, A. A cost utility analysis of treatment options for gallstone disease: Methodological issues and results. *Health Economics* (1994); **3**:157-168.

14. Creese A, Floyd K, Alban A, Guinness L. Cost-effectiveness of HIV/AIDS interventions in Africa: a systematic review of the evidence. *Lancet* (11 May 2002); **359**:1635-43.
15. Currie GR, Donaldson C, McIntosh E. Cost effectiveness analysis and the "societal approach": Should the twain ever meet? *Institute of Health Economics - Working Paper Series* (1999); **00-3**.
16. Davis A, Whinston AB. Welfare Economics and the Theory of the Second Best. *Review of Economic Studies* (1965).
17. DeSerpa AC. Microeconomic Theory. USA: Allyn and Bacon Inc., (1988).
18. Donaldson, C., Schackley, P., Abdalla, M., and Miedzybrodzka, Z. Willingness to pay for antenatal carrier screening for cystic fibrosis. *Health Economics* 4, 439-452. (1995).
19. Dorrington, R., Bourne, D., Bradshaw, D., Laubscher, R., and Timaeus, I. The impact of HIV/AIDS on adult mortality in South Africa. (2001).
20. Drummond M, O'Brien B, Stoddart G, Torrance G. Methods for the Economic Evaluation of Health Care Programmes. United Kingdom: Oxford University Press, (1997).
21. Evans D, Hurley SF. The Application of Economic Evaluation Techniques in the Health Sector: The State of the Art. *Journal of Economic Development*, Volume 7, No 3, pp 503-524
22. Faith RL, Thompson EA. A Paradox in the Theory of the Second Best. *Economic Inquiry* (April, 1981).
23. Fox-Rushby JA, Hanson K. Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis. *Health Policy and Planning* (2001); **16**:326-31.
24. Gafni A. Using willingness-to-pay as a measure of benefits: What is the relevant question to ask in the context of public decision making about health care programs. *Centre for Health Economics and Policy Analysis: Department of Clinical Epidemiology & Biostatistics; McMaster University* (1990).
25. Garber AM. Advances in Cost-Effectiveness Analysis of Health Interventions. *NBER Working Paper* (June 1999); **No. w7198**.
26. Garber AM, Phelps CE. Economic Foundations of Cost Effectiveness

Analysis. *NBER Working Paper Series* (September 1992); **No. w4164**.

27. Garber AM, Weinstein MC, Torrance GW, Kamlet MS. Chapter Two: The Theoretical Foundations of Cost Effectiveness Analysis. In Gold MR, Siegel JE, Russell LB, and Weinstein ME, editors, *Cost Effectiveness in Health and Medicine*, New York: Oxford University Press (1996).
28. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal* (13/06/2002); **321**.
29. Ginsberg G, Blau H, Kerem E, Springer C, Kerem B, Akstein E *et al*. Cost-benefit analysis of a national screening programme for cystic fibrosis in an Israeli population. *Health Economics* (1994);**5**:5-23.
30. Gous, S. J. Question No. 24 / Internal Question Paper No. 14. *South African Parliament* (30 May 2002).
31. Grant RM, Hecht M, Warmerdam M, Liu L, Liegler T, Petropoulos CJ *et al*. Time Trend in Primary HIV-1 Drug Resistance Among Recently Infected Persons. *Journal of the American Medical Association* (10 July 2002);**288**:181-8.
32. Gravelle H, Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Health Economics* (2001);**10**:587-99.
33. Grossman M. The Human Capital Model. *NBER Working Paper* (April 1999); **No. w7078**
34. Guyatt HL, Tanner M. Different approaches to modeling the cost-effectiveness of schistosomiasis control. *American Journal of Tropical Medicine and Hygiene* (1996); **55**:159-64.
35. Hammer SM, Vaida F, Bennett KK, Holohan MK, Sheiner L, Eron JJ *et al*. Dual vs Single Protease Inhibitor Therapy Following Antiretroviral Treatment Failure: A Randomized Trial. *Journal of the American Medical Association* (10 July 2002); **288**:169-80.
36. Hammit JK. Discounting Health Increments. *Journal of Health Economics* (1993); **12**:117-20.
37. Hanson K. Measuring Up: Gender, Burden of Disease, and Priority Setting Techniques in the Health Sector. *Harvard Center for Population and Development Studies - Working Paper Series* (August 1999); **99.12**.

38. Holtgrave DR, Kelly JA. Preventing HIV/AIDS among high-risk urban women: the cost-effectiveness of a behavioral group intervention. *American Journal of Public Health* (October 1996); **86** :1442-5.
39. Homedes, Nuria. The Disability-Adjusted Life Year (DALY) Definition, Measurement and Potential Use. Human Capital Development and Operations Policy Working Papers (World Bank) . 21.
40. Johannesson M, Jönsson B. Economic evaluation in health care: Is there a role for cost-benefit analysis? *Health Policy* (1991); **17**:1-23.
41. Johnson-Masotti AP, Pinkerton SD, Kelly JA, Stevenson LY. Cost-effectiveness of an HIV risk reduction intervention for adults with severe mental illness. *AIDS Care* (June 2000); **12**:321-32.
42. Kenyon, C., Skordis, J., Boule, A., and Pillay, K. Rationing HAART Part II: The need for a new system of rationing health services in the era of informational capitalism. *South African Medical Journal*. (Due for publication in January 2003)
43. Kreps DM. *A Course in Microeconomic Theory*. USA: Princeton University Press, (1990).
44. Lazaro A, Barberan R, Rubio E. Private and social time preferences for health and money: An empirical estimation. *Health Economics* (2001); **10**:351-6.
45. Lazaro A, Barberan R, Rubio E. The economic evaluation of health programmes: Why discount health consequences more than monetary consequences? *Applied Economics* (2002); **34**:339-50.
46. Lipsey and Lancaster. The General Theory of Second Best. *Rev. Econ. Stud.* (1965); **24**: 11-32.
47. Lipsey RG, Chrystal KA. *An Introduction to Positive Economics*. USA: Oxford University Press, (1995).
48. Mahmud Khan M, Ali D, Ferdousy Z, Al-Mamun A. A cost-minimisation approach to planning the geographical distribution of health facilities. *Health Policy and Planning* (2001); **16**:264-72.
49. Marseille E, Hoffman PB, Kahn JG. HIV Prevention Before HAART in Sub-Saharan Africa. *Lancet* (2001); **359**:1851-6.

50. Mechanic D. Dilemmas in rationing health care services: the case for implicit rationing. *British Medical Journal* (1995); **310**:1655-9.
51. Mechanic D. Approaches to controlling the costs of medical care: short-range and long-range alternatives. *The New England Journal of Medicine* (February 2, 1978);**298**:249-54.
52. Medicin Sans Frontieres. Affordable Antiretrovirals in South Africa: Protocol. *MSF* (2001).
53. Medicin Sans Frontieres. Providing Antiretroviral Therapy at Primary Care Clinics in Resource Poor Settings. *MSF* (2002); **Cape Town**.
54. Moatti JP. Personal Communication. (August 2000).
55. Molla MT, Wagener DK, Madans JH. Summary Measures of Population Health: Methods for Calculating Healthy Life Expectancy. *Healthy People 2010; Department of Health and Human Services* (August 2001); **21**.
56. Mooney G, Wiseman V. Burden of Disease and Priority Setting. *Health Economics* (2000);**9**:369-72.
57. Morrison GC, Gyldmark M. Appraising the use of contingent valuation. *Health Economics* (1992);**1**:233-43.
58. Murray, C. J. L. *Quantifying the burden of disease: the technical basis for disability-adjusted life years*. (1994). World Health Organisation.
59. Murray CJL, Acharya AK. Understanding DALYs. *Journal of Health Economics* (1997); **16**:703-30.
60. Murray CJL, Evans DB, Acharya A, Baltussen RMPM. Development of WHO guidelines on generalized cost-effectiveness analysis. *Health Economics* (2000);**9**:235-51.
61. Murray CJL, Lopez AD. *The Global Burden of Disease*. USA: Harvard School of Public Health on Behalf of the World Health Organisation and the World Bank, (1996).
62. Murray, C. J. L. and Salomon, J. A. Using mathematical models to evaluate global tuberculosis control strategies. Center for Population and Development Studies: Harvard University. (Downloaded from the internet on 2/19/02)

63. Olsen JA. On what basis should health be discounted? *Journal of Health Economics* (1993); **12**:39-53.
64. Olsen JA. Time preferences for health gains: An empirical investigation. *Health Economics* (1993); **2**:257-65.
65. Paalman, M., Bekedam, H., Hawken, L., and Nyheim, D. A critical review of priority setting in the health sector: the methodology of the 1993 World Development Report. *Health Policy and Planning* 13((1)), 13-31. (1998).
66. Parker M, Fox -Rushby JA. International Comparisons in Health Related Quality of Life: Acquiescence in Academia. Oxford: Rapid Communications, (1995).
67. Parsonage M, Neuburger H. Discounting and health benefits. *Health Economics* (1992);**1**:71-9.
68. Peters, D. H., Kandola, K., Elmendorf, A. E., and Chellaraj, G. Health Expenditures, Services, and Outcomes in Africa: Basic Data and Cross-National Comparisons 1990-1996. (July 1999). The World Bank. Health, Nutrition and Population Publication Series.
69. Pillay K. Personal Communication. (August 2000).
70. Pinkerton SD, Holtgrave DR. How HIV treatment advances affect the cost-effectiveness of prevention. *Medical Decision Making* (Jan-March 2000);**20**:89-94.
71. Postma MJ, Beck EJ, Hankins CA, Mandalia S, Jager JC, de Jong-van den Berg LT, Cost Effectiveness of expanded antenatal HIV testing in London. *AIDS* (20 October 2000);**14**:2383-9.
72. PulseTrack. Policy shift - more evidence on process/actors. (25 April 2002).
73. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, Meehan M O, Lutalo T, Gray R H, Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1, The Rakai Project Study Group *New England Journal of Medicine* (2000); **342**:921-9.
74. Reekie WD. Health-care options for South Africa: Lessons from the UK and the USA. South Africa: The Free Market Foundation, (August 1995).

75. Richardson, J. Cost utility analysis: What should be measured? *Social Science and Medicine* 39(1), 7-21. (1994).
76. Ried, W. Willingness to pay for diagnostic services. A new approach to modelling patient benefits in health care. *Health Economics* 3, 255-266. (1994).
77. Rosen SH. *Public Finance* (1995); Irwin McGraw Hill, USA.
78. Rosenzweig MR. Population Growth and Human Capital Investments: Theory and Evidence. *The Journal of Political Economy* (October 1990);98:S38-S70.
79. Russell, S. Ability to pay for health care: concepts and evidence. *Health Policy and Planning* 11(3), 219-237. (1996).
80. Russell, S., Fox-Rushby, J. A., and Arhin, D. Willingness and ability to pay for health care: a selection of methods and issues. *Health Policy and Planning* 10(1), 94-101. (1995).
81. Schultz LJ, Steketee RW, Chitsulo L, Wirima JJ. Anti-malarials during pregnancy: a cost-effectiveness analysis. *Bulletin of the World Health Organisation* (1995);73:207-14.
82. Sinanovic E. Economic Evaluation of Healthcare. *Course Notes*, University of Cape Town (2001).
83. Skordis J, Natrass N. Paying to waste lives: the affordability of reducing mother-to-child transmission of HIV in South Africa. *Journal of Health Economics* (2002); 21:405-21.
84. Smith RD. The relative sensitivity of willingness-to-pay and time-trade-off to changes in health status: An empirical investigation. *Health Economics* (2001);10:487-97.
85. South African Department of Health. Standard Treatment Guidelines and Essential Drugs Lists. <http://www.sadap.org.za/edl/> (1998).
86. Stiglitz J. *Economics of the Public Sector*. New York/London: WW Norton and Company, (2000).
87. The Economist. *Pocket World in Figures*. Italy: Profile Books Ltd, (2001).
88. Thompson, M. S. Willingness to pay and accept risks to cure chronic

- disease. *American Journal of Public Health* 76, 392-396. (1986).
89. Treatment Action Campaign. Home Page. <http://www.tac.org.za/home.htm> (Downloaded from the internet on 22/07/2002).
 90. UNAIDS. A Global View of HIV Infection. (2002). http://www.unaids.org/barcelona/presskit/barcelona%20report/graphics/globalview_en.pdf, UNAIDS. (Downloaded from the internet on 22/07/2002).
 91. UNAIDS Reference Group. Improved Methods and Assumptions for Estimation of the HIV/AIDS Epidemic and its Impact: Recommendations of the UNAIDS Reference Group on Estimates, Modelling and Projections. *AIDS* (2002); **16**:1-14.
 92. USAID. Prevention to Care: USAID's Interrelated Strategies for Fighting the HIV/AIDS Pandemic . *USAID Global Health: HIV/AIDS, News & Information*. (2002). USAID.
 93. van de Walle D. Assessing the Welfare Impacts of Public Spending. *WHO: Policy Research Working Paper* (October 1996);**1670**.
 94. Varian HR. *Microeconomic Analysis*. USA: W.W. Norton & Company, (1992).
 95. Viscusi WK. Chapter 7 - *The Valuation of Risks to Life and Health: Guidelines for Policy Analysis*. pp 193-210. Dordrecht: D. Reidel Publishing Company, (1986).
 96. Whiteside A, FitzSimons D. The AIDS Epidemic: Economic, Political and Security Implications. *Research Institute for the Study of Conflict and Terrorism: Conflict Studies* 251 (May 1992).
 97. WHO. Commission on Macro Economics and Health. World Health Organisation, (2001).
 98. Williams, A. Calculating the global burden of disease: Time for a strategic reappraisal? *Health Economics* 1, 1-8. (1999).
 99. Williams A. Comments on the Response by Murray and Lopez. *Health Economics* (2000); **9**:83-6.
 100. World Bank. World Development Report: Investing in Health. *World Bank: Washington D.C.* (1993).
 101. World Development Indicators Database. *South African Data Profile*. (April

2002).

102. World Health Organisation. Acquired Immunodeficiency Syndrome: Interim Proposal for a WHO Staging System for HIV Infection and Disease. *Weekly Epidemiological Record* (1990);**65**:221-4.
103. World Health Organisation. The World Health Report 2001. <http://www.who.int/wbr/> (Downloaded from the internet on 23/07/2002)
104. World Resources Institute. Global Health Patterns - A New Metric: Factoring in Sickness and Disability. World Resources Institute: Health and Environment . (Downloaded from the internet on 16/07/2002)
105. Yeni PG, Hammer SM, Carpenter CCJ, Cooper DA, Fischl MA, Gatell JM *et al.* Antiretroviral Treatment for Adult HIV Infection in 2002: Updated Recommendations of the International AIDS Society-USA Panel. *Journal of the American Medical Association* (10 July 2002); **288**:222-35.
106. Zaric GS, Barnett PG, Brandeau ML. HIV transmission and the cost-effectiveness of methadone maintenance. *American Journal of Public Health* (July 2000); **90**:1100-11.