METHODOLOGICAL ISSUES IN MEASUREMENT AND VALUATION OF DISEASE STATES: A CASE STUDY OF SCHISTOSOMIASIS MANSONI IN KENYA

By

MERCY GATHIGIA MUGO

THESIS SUBMITTED TO THE UNIVERSITY OF LONDON
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

HEALTH POLICY UNIT
DEPARTMENT OF PUBLIC HEALTH AND POLICY
LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

OCTOBER 2003

DEDICATION

To Daisy, my daughter

For all the fine things that you are: patience, courage, understanding and for encouraging Mum to work hard all the time.

To Jane, a friend and second mother to Daisy

For revealing what friendship is, for being selfless. Your deeds sustained my peace of mind to soldier on.

To Wang'ombe, my mentor

For the deep massive doses of positive encouragement all the time

ACKNOWLEDGEMENTS

I am thankful and full of gratitude to all those who helped in one way or another towards the completion of this thesis. To Dr. Julia Fox-Rushby, I am forever grateful for your patience, help and being a great teacher in this new and exciting discipline. I appreciate the help from my advisory committee; Dr. John Browne for all the kind and very encouraging words- they took me a long way and to Dr. Sandy Cairncross for your kindness and advice.

My special gratitude goes to the UNDP/World Bank/ WHO Special Programme for Research and Training in Tropical Diseases (TDR/WHO) for their financial support to undertake this training and making my dream come true. My most appreciation to Steven Wayling for his prompt support and kind consideration.

I am grateful to the University of Nairobi for granting me time off work to undertake this study. In particular, I am grateful to the Chairman of Economics Department during my study, Prof. Masai for all the support and to colleagues at the Department of Economics for their encouragement and moral support. Special thanks to Prof. Wang'ombe for introducing me to health economics and whose unwavering support and encouragement has been a great inspiration. I thank Prof. Gakuru for his support and for facilitating research assistants recruitment. To Dr. Mwaniki of MRC, KEMRI, I say thank you for your support and encouragement. To Dr. John Ouma and Dr. Curtis Kariuki of Division of Vector Borne Diseases, I am so grateful for all the consultation, advise and allowing me to use your laboratory facilities and staff at Mwea Health Centre. I am also grateful to Dr. Sullyman of LSHTM for support and encouragement.

I owe my gratitude to all those Mwea residents who devoted so much of their time and effort to respond to my questionnaires repeatedly and for their hospitality. To the chief, sub-chiefs and village in-charges who provided smooth entry into the study area, receive my heartfelt appreciation. Special thanks to John, Lillian, Mundia, and Muriithi; you were a great team to work with. Your rapport with Mwea residents made it a great joy to work

with you. Special thanks to Nyaga, Lucy, Alice, Wachira, Njeru, Muriithi, Esther and all the staff at Kimbimbi Health Centre and Nguka dispensary for their assistance with patient recruitment and screening. Without your help, fieldwork for this thesis would not have been accomplished.

On the social front I am thankful to the Taviton crew for lending their ears, support and encouragement. To Rosemary, Ana and Jonny, Teck-Peng, Matthew and Rachel, Sheere, and other members of my home group at Willy G., thank you for your friendship, love, prayers, encouragement and spiritual nourishment.

I wish to express my deepest appreciation to Jane Maina, Mike Kiarie and her family for providing home, parental care and comfort to Daisy during my studies. Many thanks and appreciation to Daisy for being the brave one in enduring my absences and her high sense of understanding while I was away studying. Special thanks to Rachael Gituku for her friendship, companionship and humour. You did a great job looking after Daisy during my fieldwork. Without the love and support of my entire family, especially the Chiuri's and Gituku's, I would not be where or who I am today. I consider myself very privileged and thank God for having you as my family. Finally, to the one that I adore, the author of my every breath, JESUS. Thank you for giving me life and guiding me through. YOU ARE MY ALL.

All the errors in this work stay with me.

ABSTRACT

Aims: To contribute to the methodological debates surrounding the measurement and valuation of disease specific health outcomes for use in economic evaluation and health care decision making and to highlight the cross-cultural relevance of application of measurement and valuation techniques in low-income countries.

Methods: Theoretical, methodological and empirical literature on measurement and valuation of HRQL with special attention to disease specific utilities was extensively reviewed. A new *S. Mansoni* HRQL measurement questionnaire was developed and three health state valuation approaches tested for application in Mwea, Kenya. Content and construct validity was assessed for the measurement and valuation approaches in addition to reliability for the VAS and TTO. The measurement and valuation samples consisted of 161 and 117 respondents respectively drawn from community members and patients.

Findings: There were knowledge gaps in the literature regarding methodological issues in measurement and valuation of disease states. The new S. Mansoni HRQL questionnaire was content and construct valid with regard to symptoms and HRQL domains. Content validity for the VAS, TTO and SG was established. The VAS was more construct-valid and reliable than the TTO. There was evidence that to make HRQL outcomes comparable across settings, cross-cultural equivalence of the measurement and valuation tools is required.

Conclusion: The new *S. Mansoni* HRQL questionnaire provides a construct-valid measure of impact of disease that takes HRQL issues into account, though it requires further testing for reliability and validity in other settings. The VAS and TTO could be used to elicit valid and reliable values for disease states, with the appropriate modifications for cultural equivalence. With further testing, the SG could also be used to elicit values, in the Kenyan and other similar settings, as its content validity was ascertained, but would require testing for construct validity and reliability.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CBA	Cost Benefit Analysis
CEA	Cost Effectiveness Analysis
CHEPA	Centre for Health Policy Analysis
CUA	Cost Utility Analysis
DALY	Disability Adjusted Life Year
DSU	Disease specific utilities
EQ-5D	European Quality of Life 5 Dimensions
EUT	Expected Utility Theory
FGD	Focus Group Discussion
GBD	Global Burden of Disease
GOK	Government of Kenya
GNP	Gross National Product
HRQL	Health Related Quality of Life
HS	Health Status
HUI	Health Utilities Index
HYE	Health Years Equivalent
IPI	Intestinal Parasitic Infestations
IWI	Intestinal Worms Infestations
KENQOL	Kenya Quality of Life
Kshs.	Kenya shillings
ME	Magnitude Estimation
MOH	Ministry of Health
NGO	Non- Governmental Organisation
NIB	National Irrigation Board
OP	Office of the President
PCD	Partnership for Child Development
PEM	Protein Energy Malnutrition
PHCC	Primary Health Care Centres
PTO	Person Trade Off

PR	Principal Researcher
PUMS	Patient Utility Measurement Scales
QALY	Quality Adjusted Life Year
QOL	Quality of Life
QWB	Quality Well Being
RS	Rating Scale
SAVE	Saved Young Life Equivalent
SSA	Sub-Saharan Africa
SG	Standard Gamble
TTO	Time Trade Off
UK	United Kingdom
UN OCHA IRIN	United Nations Office for the Co-ordination of Humanitarian Affairs Integrated Regional Information Network.
VAS	Visual Analog Scale
VN-M	Von-Neumann-Morgenstern
WHO	World Health Organisation
WHO/CTD	World Health Organization Center for Tropical Diseases
WHO/TDR	World Health Organization Tropical Diseases Research
WHOQOL	World Health Organization Quality of Life

CHAPTER 1

INTRODUCTION

This thesis endeavours to make a contribution to methodological debates surrounding the measurement and valuation of health outcomes for use in economic evaluation. It is set within a cost utility analysis (CUA) framework but focuses on outcome assessment and in particular health related quality of life rather than survival. Using a case study of *Schistosomiasis Mansoni* in Kenya, this thesis examines the suitability of existing methods for measuring the impact of treatment and valuing change to quality adjust survival. The arising methodological and empirical contributions have relevance for future research and policy making in developing countries.

Schistosomiasis Mansoni (also known as bilharzia) is the second most prevalent tropical disease after malaria (WHO, 1998) and is a leading cause of severe morbidity (WHO/CTD, 1999). *S. Mansoni* afflicts the majorities of populations in developing countries. It is estimated that 80% of all the people infected with schistosomiasis are in Sub-saharan Africa (WHO, 1998). Nevertheless, the impact of *S. Mansoni* on HRQL remains unknown, which implies that issues relating to economic efficiency of schistosomiasis control interventions relative to other disease control interventions remains unknown too.

This chapter presents the rationale for the thesis by highlighting the key knowledge gaps it addresses. The aims and objectives are then presented followed by a brief description of the study area and the population. Finally, the organisation of the rest of the thesis is described by chapter.

1.1 Rationale

The presence of any illness of the body or mind in a person can have different consequences. One consequence may be mortality and another, the experiences of specific clinical signs and symptoms from the illness depending on the seriousness of the pathology. Different views regarding the second consequence and how it impacts on the person suffering the illness have contributed to a wide range of different measures of outcomes from illness. As what is measured affects what is shown to be effective and efficient, this influences the design and implementation of interventions as well as development of instruments for measuring such outcomes.

Measures of health have many uses. McDowell and Newell (1996) suggest they can be used "to indicate the major health problems confronting society, to contribute to the process of setting policy goals and to monitor the effectiveness of medical and health care". Given concerns with efficiency of health care interventions in the face of scarce resources and unlimited health care needs, health outcomes from interventions need to be assessed comprehensively and incorporate both quantity and quality gains.

In constructing measures that combine both quality and quantity of life, it is important to explicitly distinguish between measurement and valuation of health/ disease states. Valuation allows expression of measured dissimilar 'quantities of health' to be expressed on a common unit or denominator such as the quality adjusted life year (QALY) or disability adjusted life year (DALY). This is considered vital for cross-program comparisons of results from different interventions, yet there are situations where researchers are not clear about distinctions between the two steps.

Many health professionals and health care systems base delivery of health care services on perceiving health as the absence of disease, a biomedical view that has been heavily criticised within the medical sociology (Wilkin et al. 1992; Blaxter and Paterson, 1982). Based on this view, measures of population health have included life expectancy, morbidity and mortality rates. These measures however, say nothing about the health of the surviving

population nor do they give any indication about the impact of different illnesses on quality of life of those living in them (Bowling, 1997; WHO, 1998; Patrick and Erickson, 1993).

In most developed countries there has been a shift from traditional measures of health such as life expectancy, morbidity and mortality rates to measures of health that take into account both positive and negative effects of treatment or condition (McDowell and Newell, 1996; Bowling, 2001; Bowling 1997), and expressing them as both quality and quantity gains. This shift has occurred following increases in life expectancy and the emergence of incurable chronic diseases for which treatment does not necessarily prolong life. Concerns about improving the quality and not just the quantity of life necessitate a different approach to measurement of health outcomes, beyond traditional biomedical measures.

Health effects ought to be assessed in the broadest sense possible as many health care interventions are intended to improve general health and quality of life. Developing countries like Kenya are still grappling with low life expectancies and high morbidity and mortality rates compared to developed countries. However, although acute, preventable and readily treatable diseases contribute a substantial burden (MOH, 2002), chronic conditions such as lower respiratory infections, tuberculosis (TB) and acquired human immunodeficiency syndrome (HIV/AIDS) abound and have become major health concerns that threaten to consume substantial health care resources. For the preventable and readily treatable conditions like *S. Mansoni*, health care programs and interventions may have little impact on mortality (WHO/CTD, 1999; WHO, 1998). The emerging interests in global health and global health measurement exercises as typified by the calculation DALYs in the global burden of disease (GBD) and cost-effective analysis (CEA) studies by Murray and colleagues at the WHO (Murray and Lopez, 1994; Murray and Lopez, 1996), has hastened the policy and research agenda in the area of developing and using comparable measures of health across countries. Nevertheless, the shift from traditional measures of health has not taken root in many developing countries. The DALY which has been used extensively in developing country regions, does not provide a tool for primary measurement of quantity but is a tool for estimating the burden of disease using secondary data (Fox-Rushby, 2002).

This implies that the 'amounts of health' that incorporate quality and quantity gains remain largely unknown in these regions. Also unknown are issues related to economic efficiency with which health care resources are allocated to different health care programs in developing countries.

The use of health measures that take into account both the quality and quantity of life, such as QALYs has been severely limited in developing countries, and perhaps not heard of in many countries. The few studies (Jelsma, 2002; Baltussen et al. 2002; Mahapatra et al. 2002; Sadana, 2002; Kirigia, 1994; Tan Torres, 1991) undertaken in developing countries that take into account quality of life have mainly been involved with the development and testing of particular tools. This is an indication of the likely deficiencies of resource allocation decisions taken to date using such approaches. Given that interest in use of these measures in the developing world is beginning to gain currency, there is need for careful development and testing of the applicability of these tools in settings where their actual use is limited, even though estimations for policy making are influencing health sector development loans given through the World Bank (Fox-Rushby, 2002).

In measuring the quantity of health, generic or disease specific instruments can be used. Generic instruments are broad and they allow comparisons across different diseases and populations (Patrick and Deyo, 1989; Brazier et al. 1999). However, they can be insensitive and irrelevant when applied to specific diseases and hence the usefulness of disease specific instruments (Gold et al. 1996). Disease specific instruments have greater sensitivity and they focus on relevant aspects of different conditions (Bowling 2001; McDowell and Newell, 1996; Brazier and Deverill, 1999). Where logistics allow, generic and disease specific instruments should be used alongside each other, as they appear to be compliments rather than substitutes. Once the quantification of health or disease is accomplished through measurement, the quality or worth of the measured quantity is ascertained through valuation.

Preference-based index producing HRQL instruments such as the SF-6D, 15D, EQ-5D, HUI 1-3 and the QWB are most suitable for economic analysis, because through valuation, issues of efficiency can be addressed. These instruments have scoring formula based on their specified health state classification systems, which circumvents the prospect of directly valuing health states every time an instrument is used. However, none of them has been adapted and translated for use in the Kenyan setting, which implies that primary valuation would be required.

It is vital to differentiate disease specific utilities from generic utilities. Disease specific utilities (DSU) in this thesis refer to preferences (values and utilities) for disease-specific states measured using a disease-specific measurement instrument and valued using any of the existing non-monetary valuation instruments. Generic utilities on the other hand are preferences for disease/health states measured using generic health measurement instruments and valued using any of the existing valuation instruments.

Due to the sensitivity and relevance of disease-specific measurement instruments (Bowling, 2001; McDowell and Newell, 1996; Brazier and Deverill, 1999), DSU are more likely to reflect important characteristics of the disease, be responsive to treatment effects and represent aspects of health important to the patients (Revicki et al. 1998). In addition, DSU allow for computation of outcome measures such as QALYs, suitable for use in economic evaluations of disease interventions. Because DSU incorporate patients' preferences in outcome assessment and are less crude than intermediate outcome measures, they are likely to pick up small changes in health that matter to patients (Brown, 1999). Such changes might also be missed through use of generic utilities owing to their broadness as shown in Revicki et al (1998) where HUI2 utilities were less sensitive to differences in disease severity. Hence, a DSU might pick changes within a disease state that a generic utility misses because it is not specifically targeted at the particular patient group, and the changes do not move patients between generic states. Furthermore, for some diseases like schistosomiasis mansoni, there might be very small changes in HRQL that a generic utility might not detect and hence would lead to policies that allocate no resources for its treatment and control. On the other hand, DSU would pick these small changes and

considering the huge populations afflicted by the disease, these small changes would translate into a big effect on population welfare, which would elicit quite a different policy prescription.

In a way, DSU fill gaps of utilities not picked by generic utilities due to the broad nature of health state description. Thus it would be advantageous to elicit DSU especially when it can be shown that what matters to patients due to disease is related to HRQL issues, which would make DSU comparable to generic utilities. For this reason, when assessing known patient groups, it is vital to assess DSU in order to fully capture the impact and outcomes of the disease that are usable in economic evaluations.

However, to be considered as filling gaps between generic utilities, DSU ought to be assessed on a death to complete health scale as is done for generic utilities and to include assessments of the physical function, social functioning and psychological wellbeing domains, characteristic of HRQL (Revicki et al. 1998). Otherwise, as Revicki et al (1998) observe, DSU could be constrained to a particular range of the scale thereby compromising their discriminative ability and responsiveness especially in large populations with small changes. For this reason, DSU and generic utilities could be seen as complementary rather than substitutes.

There is no consensus regarding which of the five non-monetary health state valuation instruments namely, the visual analogue scale (VAS), time trade-off (TTO), standard gamble (SG), person trade-off (PTO) and magnitude estimation (ME) is best for eliciting preferences for health /disease states. However, the VAS, TTO and the SG have been used more widely and their performance in terms of validity and reliability is better known in North American and European settings. Little about their performance is known in developing country settings.

Most HRQL measurement and valuation instruments have been developed in North America and Europe in English (Fox-Rushby and Parker, 1995). Users of these instruments

in other languages or settings would need to adapt them through translation or develop new ones where existing instruments fail to achieve equivalence through translation (Herdman et al. 1997, 1998; Guillemin et al. 1993). There have been criticisms about the relevance of translated HRQL instruments arising from the quality of translations and non-accounting for cultural differences during adaptation and translation process (Bowden and Fox-Rushby, 2003; Fox-Rushby and Parker, 1995). A review of use of nine[1] HRQL outside North America and Europe prior to 2000 found no use of preference-based index producing HRQL instruments in Africa and very limited use of the EQ-5D, HUI and 15D in East Asia and Pacific and the Americas (Bowden and Fox-Rushby, 2003), although the EQ-5D has since been used in Zimbabwe (Jelsma, 2002).

The enormity of the task of developing new generic HRQL instruments (either index or profile producing), following what would be considered good practice, in terms of time, money and expertise knowledge has been noted (Bullinger et al. 1993; Guyatt, 1995; Herdman et al. 1998; Parker and Hopwood, 2000;). The adaptation procedure is also time and resource consuming and would require the involvement of developers of the original instruments and multidisciplinary teams in both target and original cultures. This would raise methodological issues of ensuring equivalence in both the target and original cultures for results from such instruments to be comparable (Herdman et al. 1997 and 1998). These two options were considered beyond the scope of this thesis, leaving the option of using a disease specific instrument. However, as no disease specific measure of HRQL for *S. Mansoni* existed this thesis sought to develop a new measure using a clinimetric approach (Fayers and Hand, 2002) following suggestions by Guyatt (1995).

The absence of a suitable generic preference-based index producing measurement instrument in Kenya left the option of developing a disease specific instrument followed by

[1] These were 15D, Dartmouth COOP Charts, EuroQol (EQ-5D), Health Utilities index (HUI); Nottingham Health Profile (NHP), short-Form 36 (SF-36), Sickness Impact Profile (SIP); Quality of Well-Being Index (QWB), and the WHOQOL.

using the available non-monetary health state valuation instruments to obtain utility weights for QALY type computations. However, use of these valuation techniques would require consideration of their conceptual basis to ensure that the valuation tools are conceptualised and used similarly in different cultural settings. While the majority of work has questioned the equivalence of measurement tools when used in other settings (Herdman et al. 1997 and 1998; Bullinger et al. 1993; Fox-Rushby and Parker, 1994), the conceptual basis of the valuation tools is implicitly assumed to be universal in the literature. This assumption is questioned in this thesis, by assessing the concepts and key terms embodied in the instruments.

Herdman et al. (1998) suggests examining conceptual equivalence at the beginning of an adaptation process, using qualitative research to establish a working idea of what the concept means in the target culture and to facilitate deciding the content of the questionnaire. This process which is recommended for measurement instruments, could be used to ensure that the concepts embodied in health state valuation instruments are similar in both original and target culture. An understanding of the similarities and differences in perceptions of concepts in two cultures is important not only for suitable adaptation of the instruments, but also for understanding what modifications in preferences are required before cross-culture and cross population comparisons of health can be made.

Beyond ensuring conceptual and content validity of instruments, other forms of validity like construct and criterion validity need to be established. Difficulties with establishing criterion validity in the health measures relate to lack of a suitable criterion. However, construct validation, which involves testing hypothesis based on constructs drawn from theory is an on-going process. Every time an instrument is used in a new setting, its performance should be ascertained (Carmines and Zeller, 1979). This includes validity and reliability of instruments, in addition to their feasibility and acceptability. Understanding of variation in values is an issue that deserves researching in new settings as this may differ across settings and cultures and would require consideration in decision making and policy formulation.

The rationale and the argument of this thesis is that to produce valid and reliable preferences for disease states that can be used in economic evaluation of control strategies using the QALY methodology, it is first important to test the working and the suitability of the existing methods. To this end, this thesis has focused on developing a tool to measure the impact of *S. Mansoni* on HRQL and testing the applicability of existing valuation approaches in the Kenyan setting using *S. Mansoni* disease states as a case study.

1.2 Aim and Objectives

The thesis aims to contribute to the methodological debates surrounding the measurement and valuation of disease specific health outcomes for use in economic evaluation and health care decision making. It aims to highlight the cross-cultural relevance of application of measurement and valuation techniques in low-income countries. Achievement of these aims will be pursued through the following objectives.

1. To critically review the use of disease specific utilities in economic evaluation and health care decision-making.

2. To explore the debates and issues raised in developing disease specific utilities in a low-income country.

3. Using schistosomiasis as a case study, establish the relationship between symptom severity, parasitological indicators and different disease states.

4. Determine the relative reliability, validity, practicality and equivalence of current approaches to valuing disease states using schistosomiasis as a case study.

5. Consider the potential use of the disease specific utilities obtained for *Schistosomiasis Mansoni* for decision making in Kenya and elsewhere.

1.3 Geographical focus of thesis

The geographical context of this thesis is Mwea, one of the three divisions comprising Kirinyaga District in Central province in Kenya. Figure 1.1 shows the location of Kirinyaga district and Mwea division. The division covers 527 square Kilometres and is divided into five administrative locations (Tebere, Mutithi, Thiba, Nyangati and Murinduko). Thiba location is divided into three sub-locations (Nguka, Thiba and Wamumu) covering an area of 90 sq.km. and containing 4,012 households with a population of 23,707 people. My work focused on Thiba location for a community survey and Mwea division for patient data.

The population structure of Kirinyaga District is pyramidal with 58% under the age of 19 years of whom 43% are of school going age (GOK, 1997). The whole population of Mwea Division was projected to reach 145,000 by 1999, with a population density of 263 people per Km^2 (GOK, 1989). Around 80% of the population in Kirinyaga District can read and write, with literacy in the 10-30 age group over 90%. Most people (97%) in Kirinyaga are from the ethnic group of Kikuyus. The predominant occupation is farming, especially for rice in Mwea. With a history of forced settlement as tenants in small congested plots, people in Mwea are amongst the poorest in the division (GOK, 1997).

Mwea division has the highest number (n=28) of health facilities in Kirinyaga District comprising of 1 GOK health center (Kimbimbi HC), 1 Non-Govermental Organisation (NGO) hospital, 16 government dispensaries, 5 NGO dispensaries and 3 outreach centers and 2 private nursing homes (GOK, 1997). Although the infant mortality rate is lower than the national average (61/1000 compared with 21/1000), GOK (1997) notes that infant mortality in Mwea could be higher than other divisions due to higher incidences of water borne diseases. The main cause of infant mortality is diarrhoea and the top six diseases in Kirinyaga district are malaria, respiratory tract infections, skin diseases (including ulcers), intestinal worms, diarrhoea and pneumonia (GOK, 1997).

Schistosomiasis Mansoni is parasitic intestinal helminthic infection. Lack of good sanitary and hygienic practice leads to contamination of the environment, which aggravates the transmission of the eggs of the schistosomes upon contact with human skin. The life cycle of the parasite has two hosts, a freshwater snail and a human (WHO/CTD, 1999; Stephenson and Holland, 1987). Pathology in humans is provoked by the eggs which remain trapped in the tissues (while others are excreted), leading to the formation of granulomas. It is the eggs and not the worm which cause damage to the intestine, bladder, liver and spleen (WHO/CTD, 1999). Prevalence and intensity of *S. Mansoni* is highly age-dependent (Warren et al. 1993) with much of the burden falling on school-age children. Tables A1.1 and A1.2 indicate that during 1984-1995, between 6-33% of all schistosomiasis cases in the country were found in Central Province out of which roughly 42.2% were in Kirinyaga district. Mwea division is endemic for *S. Mansoni* (Muthami et al. 1995) largely due to the water resource development in Mwea Irrigation Scheme.

Figure 1.1 Map of Kenya: Location of Kirinyaga District and Mwea Division

Source: GOK, 1997

1.4 Organisation of the thesis

This thesis is divided into a further 9 chapters. Chapter 2 provides the theoretical framework underlying the measurement and valuation of health outcomes for use in economic evaluation. The chapter starts by examining the welfare theory as the basis for cost benefit analysis (CBA) and cost effectiveness analysis (CEA). It then looks at the consumer theories under certainty that have been forwarded as a possible theoretical basis for valuation techniques such as the time trade off (TTO). Consumer theories under uncertainty outline the Von Neumann-Morgenstern expected utility theory that underlies the standard gamble (SG) technique followed by the alternatives of regret, disappointment and prospect theories.

Chapter 3 highlights the methodological and conceptual issues in measurement and valuation of health related quality of life in general, with special attention on disease specific outcomes. The chapter identifies methodological gaps that set the scope of the empirical work covered in chapters five to eight.

Chapter 4 presents a critical review of the current state of the art in eliciting disease specific utilities. The chapter focuses on methodological and empirical issues relating to disease specific utilities, with particular consideration of eliciting disease specific utilities for economic evaluation in developing countries. The chapter also determines how issues of validity, reliability and practicality as well as equivalence of instruments have or have not been addressed in relation to disease specific utilities. In looking at the extent of use of instruments in other settings, the chapter focuses on the extent to which issues of cross-cultural adaptation have been addressed. The chapter further helps in identifying methodological and empirical gaps in knowledge.

Chapter 5 is devoted to the development of a tool to measure the impact of *S. Mansoni* on HRQL. After presenting approaches to questionnaire development, the chapter examines literature on *S. Mansoni* and its impact on health status. This sets the basis for construction

of a long form questionnaire that through patient and expert opinions is modified to a shorter form questionnaire consisting of symptoms, HRQL domains and measurement of infection intensity. The methods followed in the development of the tool are intended to maximise the content validity of the instrument.

In chapter 6, results from use of the questionnaire developed in chapter 5 amongst *S. Mansoni* patients and community members are presented. The chapter focuses on establishing the validity of the measurement of disease specific states amongst the Kikuyu in Kenya. It also assesses the impact of *S. Mansoni* in 3 ways: symptoms, HRQL domains and infection intensity. It also attempts to establish the relation between the symptoms, HRQL domains and infection intensity, thereby allowing assessment of construct validity of the *S. Mansoni* HRQL questionnaire.

Chapter 7 concerns the development of an approach to valuing disease states in Kenya. The chapter begins with a reasoned presentation of the construction and choice of disease states to be valued in the empirical study based on findings from chapters 5 and 6. Approaches to pre-testing and justification of the choice of valuation techniques to be used in the valuation study are described, followed by a presentation of results. The discussion considers methods used in assessing content validity of the valuation approaches and the extent to which they could be considered to have attained content validity in the Kenyan setting. It also discusses whether any of the valuation approaches should be used in Kenya.

In chapter 8, results from application of two valuation techniques, the VAS and TTO, in rural Kenya are presented. The results primarily focus on the validity, reliability and practicality of the two approaches, with some consideration of the factors affecting variation in values across and within disease states. The discussion focuses on the level of performance of both the VAS and TTO in the Kenyan setting, exploring what might have influenced their performance.

Chapter 9 draws the findings of the thesis together, and focuses on five questions. It first explores whether the new approach to assessing impact of *S. Mansoni* on HRQL is better

than existing outcome measures. The issue of which valuation instrument has the strongest base for application in Kenya is then examined, followed by an examination of how well the TTO and VAS can cross cultures. The fourth question debates whether the values elicited for disease states in the thesis were representative of the impact of *S. Mansoni* on HRQL. Finally an illustration, using values for disease states obtained in the thesis, of the potential policy implications of using a CUA in an economic evaluation of a *S. Mansoni* intervention is presented and discussed.

Chapter 10 presents the conclusions. It begins with a summary of key findings by chapter followed by an outline of the principal methodological and empirical contributions to knowledge. Lastly, suggestions for future research are made.

CHAPTER 2

THEORETICAL FRAMEWORK OF THE THESIS

2.0 BACKGROUND

Economics is concerned with the allocation of scarce resources. The aim of economic evaluation of health care programs is to serve as an aid to decisions and to affect policy making. These decisions concern allocation of resources to health care programs to improve efficiency in resource use. Two major techniques that have been used to guide resource allocation decisions in health care programs are cost benefit analysis (CBA) and cost effectiveness analysis (CEA), with cost utility analysis (CUA) a special case of CEA. Distinctions between these techniques are based on how the outcomes are measured, the policy question each technique attempts to address and different notions of efficiency.

CBA measures both costs and outcomes or effects in monetary terms, takes a societal perspective and seeks to answer the question, "is it worth achieving this goal?" It involves interpersonal comparisons of preferences and deals with questions of allocative efficiency in resource use (Donaldson, 1998). CEA on the other hand, expresses outcomes in physical units[2]. These types of outcome measures are very restrictive with respect to extent of program or treatment effectiveness comparisons and fail to take into account effects on both quality and quantity improvements that result from an intervention. CEA seeks to answer the question, "given that it has been decided that a certain goal (e.g. health improvement through morbidity reduction of a given disease) is to be achieved, which is the least cost way of doing so (e.g. various strategies or interventions that achieves the goal)?" (Donaldson, 1998). CEA always involves comparison of at least two options with the same goal (or the same budget) and as no interpersonal utility comparisons are made CEA deals with issues of technical efficiency. Cost utility analysis, a special case of CEA, overcomes the problem of comparability of program and treatment effectiveness. In CUA,

[2] These include intermediate effectiveness measures and measures based on types of events. Examples include, cure rates, reduction in infection intensity, life years saved, lives saved, etc.

QALYs which incorporate the effects on both quality and quantity of life are used as the effectiveness measures, thereby enabling comparisons of programs producing different types of health outcomes (Birch and Gafni, 1992). Since, CUA allows inter program comparisons, it can potentially answer the question of "is it worth it" to use resources in one program and not the other? This implies that CUA can potentially address issues of economic efficiency given a particular good, say health improvement.

In order to make decisions, the policy maker may face different choice situations where the outcomes of each choice depend on the state of the world that prevails. How 'should' they choose what to do in the absence of free markets? One approach is to consider what consumers would do if a free market existed. Consumer theories under conditions of certainty as well as uncertainty attempt to explain how individuals ought to behave when confronted with various choice situations. The choices that consumers make are a reflection of their value functions among various levels of a commodity (say health states) as well as across commodities (say health versus environmental goods in case of CBA). Therefore, consumer theories under conditions of certainty and uncertainty can be used to predict how for example different states of health would be valued. This information is useful in assessment of outcomes for use in CEA/CUA, the results of which feeds into resource allocation issues that can be addressed within the framework of the welfare theory. However, as Gold et al (1996) state, it is only recently have economists sought to graft CEA to theoretical roots in welfare economics.

This chapter aims to present the theoretical framework underlying assessment of health outcomes for use in economic evaluation. Welfare theory is reviewed followed by the various consumer theories, with and without certainty, that attempt to explain how individuals ought to make choices that are consistent with the principles of welfare theory. The review leads to justifying choice of which different techniques currently in use could be used to measure and value health outcomes.

2.1 WELFARE THEORY

Welfare economics addresses normative questions because it embodies certain value judgements (Drummond et al. 1997). Two key value judgements are the proposition that social welfare should comprise the sum of individual's welfare and individuals should be considered the best source of information on their own welfare, i.e. consumer sovereignty. It is the individuals themselves who decide whether they are better off or worse off with a change (Johannesson, 1996). Welfare theory relies upon an utilitarian conception of justice (greatest good for the greatest number) as opposed to the egalitarian approach (equal rights, benefits and opportunities for everybody) as encapsulated in Rawl's theory of justice (Swenson, 1992; Rawles, 1989). Welfare theory also assumes that resource allocation occurs within perfectly competitive markets where equilibrium is achieved and that the current income distribution is appropriate (Drummond et al. 1997). These propositions lay the basis for the Pareto Principle (Drummond et al. 1997), a fundamental value judgement made in welfare economics. The Pareto principle states that a change is desirable if it makes some individuals better off without making some other individuals worse off. The Pareto principle is usually coupled with the consumer sovereignty principle[3].

According to the first theorem of welfare economics, a competitive general equilibrium is under certain assumptions Pareto optimal so that with a set of prices all markets clear, i.e. there is no excess demand or supply (Johannesson, 1996). The second theorem of welfare economics states that under certain assumptions it is possible to attain any Pareto optimal situation as a result of a general competitive equilibrium given the distribution of income (Johannesson, 1996). To attain Pareto optimal situation three sets of conditions must hold. The first is the condition of efficient exchange where the marginal rate of substitution between two consumer goods must be the same for all households that consume that good. The second is the condition of efficient allocation of factors where the marginal rate of transformation between the two goods is the same for all production factors. The third is the condition of efficient output choice where the marginal rate of substitution between two

[3] In consumption of some health care services, the assumption of consumer sovereignty is violated due to existence of market imperfections and the special characteristics of health care demand such as externalities, asymmetry of information and existence of supplier induced demand (the principal-agency relationship), and uncertainty.

goods has to equal the marginal rate of transformation between the two goods (Johannassen, 1996). The Pareto principle therefore implies technical and allocative efficiency in that there is efficiency in both consumption and production as the marginal rate of substitution in consumption and the marginal rate of substitution in production are equal (Dinwiddy and Teal, 1996).

The Pareto principle says nothing about the distribution of goods. This is to say that a Pareto optimal situation may be one where goods are highly unequally distributed in the economy ((Dinwiddy and Teal, 1996). There are a number of cases where the market fails to achieve an efficient outcome (Johannassen, 1996), such as in non-traded and public goods due to market distortions, calling for public intervention in the market. Public interventions lead to changes where some individuals gain and some individuals lose, and in those cases the Pareto principle cannot be used to determine whether a change should be carried out or not (Johannassen, 1996). Because the majority of policies produce both gainers and losers, the Pareto principle is of little practical use (Johannassen, 1996).

The limitations of Pareto Principle can be overcome theoretically in terms of the compensation criteria put forward by Hicks and Kaldor in the 1930s (Dinwiddy and Teal, 1996), which were based on the idea of a potential Pareto improvement. This criterion, known as the Kaldor-Hicks compensation principle states that "if an economic policy has the consequence of making one set of people better off and another set worse off, a potential Pareto improvement can be said to have occurred if the gainers could compensate the losers and still benefit from the change" (Dinwiddy and Teal, 1996). According to Kaldor's compensation test, a change is desirable if gainers can hypothetically compensate losers and still be better off than without the change while according to Hicks, a change is desirable if losers cannot hypothetically "bribe" the gainers and still be better off than with the change. Unlike the Pareto principle, the compensation principle does not require the actual payment of compensation. The rationale for the compensation test is that if benefits exceed costs, then if compensation were costless, it would be possible to redistribute income or re-allocate the existing bundle of goods leading to change in patterns of

production and distribution, so as to achieve an actual Pareto improvement (Johannassen, 1996) and hence net gains to society.

The Potential Pareto Improvement is criticized for being only 'potential' as no actual compensation is paid thereby weakening the claims of improvement in social welfare (Johannesson, 1996). Also compensation can attract administrative costs (Dinwiddy and Teal, 1996).

The welfare theory through the potential Pareto improvement criteria provides us with a tool to decide whether resource allocation decisions are efficient. These decisions essentially rely on choices that individuals make with respect to different goods and services. The Pareto conditions depend on both producer and consumer theory. As this thesis is only considering valuation of disease states, only the consumer theories will be considered. Consumer or demand theories help to explain how consumers ought to make choices between different consumption bundles, and are reviewed next.

2.2 CONSUMER THEORIES

The notion of choice lies at the core of theoretical explanations of consumer behaviour. Consumer theory assumes the consumer to be rational such that given her income and market prices, she spends her income to attain the highest possible satisfaction or utility[4]. Full knowledge of all the information relevant to her decision is assumed (Koutsoyiannis, 1987). To make these utility maximising decisions, the consumer must be able to compare utilities from different bundles of goods and services. Consumer preferences give an indication of satisfaction or the utility that consumers derive from consuming certain goods. The more preferred the good the more utility the consumer attains from its consumption. Preferences are the fundamental description useful for analysing choice (Varian, 1990) and they can be described using utilities or values (Drummond et al. 1997).

[4] Referred to as the axiom of utility maximisation.

It is desirable that preferences for health states be obtained in a choice based context because in making choices consumer's show their preferences and the values they attach to different health or disease states (Brazier et al. 1999b). Also consumers' utility functions can be discerned using their preferences. For purposes of interpersonal preference comparisons, elicited preferences should be at least on an interval or ratio scale.

In the next section, consumer theory under certainty is reviewed in the light of its appropriateness in describing consumer behaviour in making choices as well as representing and measuring preferences for use in health care decision making.

2.2.1 Consumer theory under certainty

Measured preferences can be ordinal or cardinal. In the cardinalist approach a number indicating quantity represents preferences. In the ordinalist approach a number indicating the ordered position represents preferences. Cardinal utility under certainty is presented first followed by the indifference curves approach and revealed preference approach representing ordinal utility under certainty.

2.2.1.1 Cardinalist utility theory

Nineteenth century economists, in developing consumer theory, assumed the existence of a cardinal utility function that represented consumers satisfaction for various bundles of commodities received with certainty. Philosophers of the time also used the concept as the foundation for utilitarian ethics in which utilities among individuals were compared and aggregated to decide on the socially optimal policy (Varian, 1990; Sen, 1982).

The Marshallian demand theory uses cardinal utility and enables development of a measure of consumer surplus that can be used to approximate the intensity of utility (Laidler, 1981). The theory is axiomatic and is based on the following assumptions (Koutsoyiannis, 1987).

1. The consumer is rational and aims at maximising utility subject to his given income constraint.
2. Cardinal utility, i.e. utility of each commodity is measurable and the most convenient measure is money.
3. Constant marginal utility of money. This is necessary if money is to be used as the measure of utility.
4. Diminishing marginal utility, i.e. utility gained from successive units of a commodity diminish.
5. The total utility of a 'basket of goods' depends on the quantities of the individual commodities; $U= f(x_1,x_2,.....x_n)$. This implies that utility gained from consuming any quantity of x_1 is independent of the quantity of x_2 consumed and vice versa.

Given these assumptions, a demand curve could be derived based on the axiom of diminishing marginal utility, which specifies consumer's equilibrium as the point of equality of the price and marginal utility of the good.

However, the assumptions of certainty, constant money income and independence in utility gained from consuming different goods[5], place a limitation on measurement of consumer surplus. Measurement of consumer surplus depends on compensating and equivalent variation which in the Marshallian formulation yields that same answer no matter what the measure (Laidler, 1981), thereby making this theory of little use in health care decision making. As Glahe and Lee (1981) note, in addition to the assumption of cardinal utility being questionable and unrealistic, *"almost all the conclusions that follow from the assumption of cardinal utility can be derived from ordinal utility assumptions and the ordinal approach provides some insights obscured by the cardinal approach"* (p 108).

[5] Assumes that goods are neither substitutes nor compliments.

2.2.1.2 Ordinalist utility theory

The ordinal utility approach has dominated the analysis of consumer behaviour in economics. The two main variants of this approach are the indifference curves approach and revealed preference theory. These two theories use ranking of preferences guided by a set of axioms to infer the consumer's utility function. They are reviewed below.

2.2.1.2.1 Indifference curves

In considering consumer's preferences, three preference relations have often been used. These are strictly preferred ($\succ$), weakly preferred ($\succsim$) and indifferent ($\sim$). It is assumed that given any two consumption bundles, the consumer can rank them as to their desirability using one of the preference relations (Glahe and Lee 1981; Varian, 1990) and guided by the axioms of the theory.

The axiom of completeness implies that any two bundles can be compared and the consumer is able to express a preference or indifference between them. The axiom of reflexivity ensures that every bundle belongs to at least one indifference set, namely that containing it if no other. The axiom of transitivity implies that if X is at least as good as Y and Y is at least as good as Z, then X is at least as good as Z. Intuitively, this is a consistency requirement on the consumer and it also ensures that no bundle can belong to more than one indifference set.

These three axioms can be used in the construction of indifference curves which show all the bundles that the consumer perceives as being indifferent to each other (Varian, 1990). Hence, all bundles that are ranked as having the same utility belong to a particular indifference set which can be represented by an indifference curve. It is assumed that consumers are able to rank all goods from the least preferred to the most preferred such that the bundles in the most preferred are represented on a higher indifference curve.

A set of further assumptions are used to define the features of a well-behaved indifference curves. The assumption of monotonicity establishes a relationship between the quantities of goods in a bundle and its place in the preference ordering in that the more of each good it contains the better. The assumption implies that more is better, hence assumes *goods* not *bads* (Glahe and Lee, 1981) and that the indifference curves have a negative slope. The assumption of continuity and preference for averages rather than extremes implies that an indifference set is a continuous surface with no gaps or breaks and convexity of the indifference curves (Glahe and Lee, 1981). Strict convexity ensures that the indifference curves have no flat spots and that it is well rounded. This becomes important in ensuring a unique utility maximisation position and not a set of infinitely many points. The curvature also implies a common feature of consumer preferences. This is the fact that the smaller the amount of good 1 held and the larger the amount of good 2 held, the more valuable are the marginal changes to the consumer in good 1 relative to marginal changes in good 2 (Glahe and Lee, 1981). The consumer will be willing to give up larger amounts of the good in abundance to obtain a unit of the other good, whose marginal value is higher.

Given well behaved indifference curves and a budget constraint (specified by the prices and consumer's income), a change in price of one good holding money income and the price of the other good constant, results in an income and a substitution effect. Income effects shift the consumer to a higher indifference curve (higher utility). The consumer can however be compensated (reduction in money income) to restore him to his original utility or alternatively an equivalent variation (increase in money income) to put him on the higher utility. These two concepts measure the consumer's surplus. Compensating variation and equivalent variation have been used to measure the monetary value (willingness to pay or willingness to sacrifice) associated with movements from one utility level to another. Note, however, that the level of utility itself cannot be measured, since it remains ordinal. The approach of willingness to pay has been used to elicit values attached to improvements in health status.

Although its assumptions are less stringent than for the cardinal utility approach, the indifference curves theory has been criticised for retaining most of the weaknesses of the

cardinalist school with the strong assumption of rationality[6] and the concept of marginal utility implicit in the definition of marginal rate of substitution (Koutsoyiannis, 1987). The theory also assumes the existence and the convexity of the indifference curves but does not establish their existence or their shape. These weaknesses imply that the theory gives no indication of how much satisfaction a bundle of goods provides one individual relative to another individual[7] due to arbitrary assignment of numerical values to indifference curves. In addition, the axioms of the indifference curves approach imply certainty. As this is not a common feature in health care demand, the theory lacks a sufficient interface of risk that characterises decisions in the real world. Hence, the indifference theory analysis proves less useful where interpersonal utility comparisons and statistical manipulation for decision making in health care is called for.

2.2.1.2.2 Revealed preference approach

The revealed preference hypothesis put forward by Paul Samuelson in 1938 has gradually taken hold of choice theory in general and demand theory in particular (Sen, 1982). The theory is based on the weak axiom of revealed preference, which guarantees consistency and prevents the violation of transitivity. The approach also assumes rationality in that the consumer prefers bundles with more goods (Koutsoyiannis, 1987). This approach involves observing people making choices and inferring preferences from those choices. The revealed preference axiom makes it possible to establish the law of demand directly without resorting to restrictive assumptions and it also establishes the existence and convexity of the indifference curves, although they are redundant in deriving the law of demand (Koutsoyiannis, 1987). The theory introduced the notion of using the budget line to understand demand, such that preferences are inferred from observing choices made by consumers between bundles of goods for a given budget. The approach of revealed

[6] It is questionable whether the consumer is able to order his preferences as precisely and as rationally as the theory implies. The preferences of the consumer could change continuously under the influence of various factors, so that any ordering of these preferences, even if possible, should be considered as valid for the very short run. The theory does not analyse the effects of advertising, past behaviour (habit persistence), of stocks of the good, interdependence of preferences of consumers, which lead to behaviour that would be considered irrational, and hence ruled out by theory (Koutsoyiannis, 1987: 28).

[7] This means that preferences are measured on an ordinal scale and therefore marginal utility analysis and interpersonal utility comparisons cannot be done.

preference has been used in studying preferences revealed by both market and non-market behaviour such as government decisions and choices of public bodies among others[8].

Although the revealed preference hypothesis was a major advancement over the classical cardinal and indifference curves approaches to derivation of demand, it has some limitations. Like the two earlier theories, it assumes certainty in choice, rationality of the consumer, that individuals are always able to express preferences over bundles of goods and does not allow for indifference between any two bundles of goods. Although the theory does not require the use of the concept of utility (Koutsoyiannis, 1987), it is implicit in its assumptions that utility can be measured ordinally. Sen (1982) demonstrated the weaknesses of these assumptions and noted that individual preferences are not bounded by his preferences only since man is a social animal. The inherent conflict between individual rationality and social optimality may require an understanding of cultural orientation of behaviour in relation to choice and decision making.

The theories reviewed above assume certainty, which puts a limitation on their usefulness in health care decision making. As Laidler (1981) notes that, *"so long as we deal with questions of choice under conditions of certainty, the ordinal utility assumption suffices as a basis for consumer theory, but, ...a cardinal function is extremely useful in dealing with choice in conditions of risk"* as it allows measurement of the strength of preferences on an interval or ratio scale and interpersonal preference comparisons. The majority of decisions in health and health care, such as choice of treatment options involve risks as the consequences are uncertain. A review of how individuals ought to make decisions under conditions of risk and uncertainty is presented next.

[8] Note that in health care it is not always possible to observe revealed preferences due to the special characteristics of demand for health care, i.e., information asymmetry, existence of supplier induced demand for health care, uncertainty, and other market distortions and imperfections in the market for health care.

2.2.2 Consumer theory under uncertainty

Most decisions in health care are accompanied by varying degrees of risk and uncertainty (Smith, 1996). For example, in clinical decision making, treatment options may carry some risk of mortality and therefore choice involves gambling over outcomes of the chosen option. This characteristic of health care requires that elicitation of preferences over available treatment options as well as the desirability of different health states incorporate risk. The Von Neumann-Morgenstern (vN-M) expected utility theory (EUT) underlies most contemporary work on risk and decision-making (O'Brien, 1990). However, the theory has been challenged with a number of studies (Loomes and Sugden, 1982, 1987; Kahneman and Tversky, 1979) demonstrating axiom violations and putting forward alternative theories.

In the next section, the EUT theory is reviewed together with evidence on axiom violations before presenting the alternative theories attempting to explain the weaknesses of the EUT as a descriptive and prescriptive theory of decision making under conditions of risk and uncertainty. The alternative theories include the prospect theory, regret and disappointment theory.

2.2.2.1 von Neumann-Morgenstern expected utility theory

It was in the search for an answer to the St. Petersburg paradox[9], illustrated by Bernoulli, that expected utility theory was formulated (Biswas, 1997). EUT, which has dominated the field of decision making and analysis of economic behaviour under risk and uncertainty for over half a century (Krabbe, 1998), was developed by the mathematician John Von Neumann and the economist Oscar Morgenstern in 1944 (Drummond et al. 1997). The theory has been referred to as the cornerstone of neo-classical demand analysis (McGuire et al, 1988). The theory is normative or prescriptive in that it prescribes how a rational

[9] This was a game in which the gambler is paid 2^n dollars if the head appears on the n^{th} toss of a coin. The mathematical expectation of the game ($\Sigma(2^n)(1/2^n)$) is infinite and therefore anybody should be prepared to pay any arbitrarily large amount of money to play. Bernoulli suggested that people chose the gamble that maximises the expected utility from the gambles rather than expected return from the gamble.

individual ought to make decisions when faced with uncertain outcomes to increase her/his welfare in the most efficient way (Von Neumann and Morgenstern, 1953; O'Brien, 1990). It assumes that individuals wish to be rational with respect to their objectives and preferences and that no one would knowingly violate the axioms of the theory (O'Brien, 1990). The theory is also a descriptive model that states that observed behaviour could be described and explained as if expected utility were being maximised (O'Brien, 1990).

Von Neumann and Morgenstern, using a set of axioms, defined what they meant by rational behaviour under uncertainty. The axioms of EUT as stated by Von Neumann and Morgenstern (1953) are reproduced below. Consider a system U of entities $u, v, w...$, where U is a system of (abstract) utilities. In U a relation is given as $u>v$ and for any number α $(0<\alpha<1)$ an operation $\alpha u + (1-\alpha) v = w$, where α is probability. These concepts satisfy the following axioms.

3Aa: ***$u>v$ is a complete ordering of U.*** This means that for any u, v one and only one of the following relation holds: $u<v$, $u=v$, $u>v$.

3Ab: $u>v$, $v>w$ imply $u>w$.

3B: ***Ordering and combining.***

3Ba: $u<v$ implies that $u< \alpha u + (1-\alpha) v$.

3Bb: $u>v$ implies that $u> \alpha u + (1-\alpha) v$.

3Bc: $u<w<v$ implies the existence of an α with $\alpha u + (1-\alpha) v <w$.

3Bd: $u>w>v$ implies the existence of an α with $\alpha u + (1-\alpha) v >w$.

3C: ***Algebra of combining.***

3Ca: $\alpha u + (1-\alpha) v = (1-\alpha) v + \alpha u$.

3Cb: $\alpha(\beta u + (1-\beta) v) + (1-\alpha) v = \gamma u + (1-\gamma) v$; where β and γ are probabilities and $\gamma = \alpha\beta$.

3A (a, b) is a statement of the completeness of the system of individual's preferences, representing the axiom of transitivity. Transitivity of preference is a plausible and generally

acceptable property (Bernard, 1986). 3B (a, b, c) represents the axiom of independence and continuity, which excludes any kind of complementarity and substitutability and ensures conformity of preferences. 3C (a, b) represents the axiom of complexity, which requires that individuals obey the laws of compound probability (O'Brien, 1990). Implicit in the formulation of these axioms is the assumption of monotonicity, i.e. desire for higher probability of success (O'Brien, 1990).

The major contribution of vN-M was the development of the neo-classical economic theory into the realm of decisions involving risk and the introduction of new form of cardinality in utility assessment (O'Brien, 1990). The axioms of EUT are "*.....sufficient to prove that there exists a utility index, unique up to positive linear transformations, so that computing expected utilities will yield a preference ordering among lotteries in accordance with the axioms*" (Schoemaker, 1981 *cited in* O'Brien, 1990). Because of their interval scale properties, the cardinal utilities can be averaged across individuals and used in aggregates as basis of social policy (Sloan, 1996), as it allows analysis of marginal changes.

The axioms of EUT have been criticised. Hey (1979) notes that the axiom of continuity is contentious in that in some situations the individual may be unable to make a choice with respect to the options available[10]. The assumption of independence implies that choices that people plan to make in one state of nature should be independent from the choices that they plan to make in other states of nature[11] (Varian, 1990; Cohen, 1996). This requires separability of preferences across different events (Fishburn and Wakker, 1995) and denotes the exclusion of any effect of complementarity or substitutability (Fishburn and Wakker, 1995; Von Neumann and Morgenstern 1953). The assumption would therefore imply that states of nature have no effect on choices and that utility from consumption of a good is independent of other goods. In terms of valuation of health outcomes, this assumption implies that the value assigned to 'quality of a health outcome' should be

[10] For example due to nature of health care and market imperfections, consumers have to rely on health providers in making their consumption decisions, or they are too ill to do so, or could be frightened by choices that could involve their own death.
[11] Different outcomes must be consumed separately (e.g quality of health and quantity of health do not affect each other), which implies additivity across different contingent consumption bundles.

independent of 'duration of the outcome' and vice versa which is an unrealistic and restrictive assumption (Kirsch and McGuire, 2000; Bala et al. 1998; Brazier et al. 1999b).

The EUT assumes that the consumer is sufficiently able, willing and knowledgeable (in terms of choices, states of the world, final consequences, probabilities and utility assessment) to make the relevant choices, in his goal of maximizing utility (McGuire et al, 1988). This assumption is rather unrealistic in the context of health care, where more often than not the patient entirely relies on the physician to make the decisions. McGuire et al (1988) note that in the consumption of health care much of consumer's sovereignty assumed by EUT is lost or eroded due to the patient-doctor agency relationship characteristic of demand for health care. Another underlying assumption of the EUT is that the relevant utility bearing characteristics are consequences or outcomes of the final states, and not processes (McGuire et al, 1988). This assumption ignores the behavioural processes that are involved in making choices, which may be influenced by other psychological factors. Although the EUT addresses the problem of uncertainty in health care decision making, its usefulness is limited in addressing other risk bearing/avoidance characteristics of health care such as information asymmetry and process utility (McGuire et al. 1988). Hence, "*in terms of a spectrum which stretches from commodities where expected utility theory fits very well to those where it fits badly, health care is at the latter end*" (McGuire et al. 1988). This implies that conventional demand theories may have limited application in health care.

The inadequacies of the EUT have been demonstrated by various paradoxes, among them the Allais paradox[12], which demonstrate violation of the axioms of EUT (Bernard, 1986; Biswas, 1997). Sugden (1989) notes that a large amount of experimental evidence has been published revealing consistent violations of the axioms of EUT. These violations suggested that the theory had major weaknesses as a predictive theory of choice under risk and

[12] In this paradox, subjects are presented with two lotteries and asked to choose one. In lottery 1 the choice is between (a) receiving 4 million with certainty or (b) receiving 10 million with probability of 0.25, 4 million with probability 0,74 and 0 million with probability 0.01. The second lottery involves (c) receiving 4 million with probability 0.26 and 0 million with probability 0.74 or (d) 10 million with probability 0.25 and 0 million with probability 0.75. A common response pattern is to choose (a) over (b) in lottery one and (d) over (c) in lottery two. However, the first preference implies $0.26\ u(4) > 0.25\ u(10)$ while the second preference implies $0.26\ u(4) < 0.25\ u(10)$. This shows preference reversals and violation of expected utility theory axioms.

uncertainty (Kahneman and Tversky, 1979; Loomes and Sugden, 1982). Alternative theories (the prospect, regret and disappointment) have been formulated to deal with the problems of decision making under uncertainty. Proponents of these theories consider them the general case with EUT only being a special case within their formulation.

In the next section, prospect theory is reviewed, followed by regret and disappointment theories.

2.2.2.2 Prospect Theory

This theory was put forward by Kahneman and Tversky in 1979. The authors presented several choice problems involving gambles to university staff and students in Israel, Michigan and Stockholm. The problems involved: choosing between certain outcomes and gambles; outcomes formulated in terms of gains and losses; compound probabilities; and in some cases involved very small probabilities of a large gain or loss (Kahneman and Tversky, 1979). The authors observed violations of the axiom of transitivity and consistency. The violations were explained in terms of the 'certainty effect' where people chose certain outcomes where the alternative was a gamble and where both choices were gambles, they choose the one with largest gain. They also found that preferences between gains were mirror images of preferences for losses and termed this as a 'reflection effect', which revealed an over-weighting of certainty in that there was risk aversion on the domain of gains and risk seeking in the domain of losses. Additionally they found that people disregarded shared components amongst alternatives and focused on components that distinguish them in making choices and termed this as 'isolation effect'. This decomposition may produce inconsistent responses because it can happen in a number of different ways. They also noted that carriers of value or utility were changes of wealth rather than the final asset positions that include current wealth and termed this as the cornerstone of their theory.

Prospect theory has been proposed as a model of choice and a useful framework for the descriptive analysis of choice under risk. The theory has two themes; one, the editing

operations that determine how prospects are perceived; two, judgemental principles that govern the evaluation of gains and losses and the weighting of uncertain outcomes (Kahneman and Tversky, 1979). The editing phase also referred to as framing is an initial screening of the options at hand with the function of re-organizing the problems to make the evaluation and choice simpler (O'Brien, 1990). This approach has conceptual appeal of how individuals process information. They edit the problem and then evaluate the simplified task. This theory rejects the idea of people being 'super-rational' data processing machines (Kahneman and Tversky, 1979). Since our ability to process information is limited, the simplification of problems may well lead to intransitive and 'sub-optimal' choice.

Prospect theory assumes that values are attached to changes rather than to final states and probabilities are replaced by decision weights. However, the value of a particular change is not independent of the initial asset position as this serves as a reference point and the magnitude of change is evaluated from this reference point (Kahneman and Tversky, 1979; O'Brien, 1990) and coded as gains or losses.

The authors hypothesize that the value function for changes of wealth is normally concave above the reference point and often convex below it. Hence, the marginal value of both gains and losses generally decrease with their magnitude (Kahneman and Tversky, 1979). A salient characteristic of attitudes to changes in welfare is that losses loom larger than gains (Kahneman and Tversky, 1979). Noting that, "*the aggravation that one experiences in losing a sum of money appears to be greater than the pleasure associated with gaining the same amount*", the authors demonstrated that the value function for losses is steeper than the value function for gains, (Kahneman and Tversky, 1979). Decision weights are lower than the corresponding probabilities, except in the range of low probabilities. Kahneman and Tversky (1979) noted that over-weighting of low probabilities might contribute to the attractiveness of both insurance and gambling.

The framing and reference effects observed by Kahneman and Tversky in money gambles have also been found in studies of medical decision-making and patient preferences (O'Brien, 1990). In a study by Eraker and Sox (1981, *cited in* O'Brien, 1990) respondents consistently chose the certain option when presented with lotteries whose expected outcome was same as the certain outcome, thereby averting risk as demonstrated by the 'certainty effect'. Individuals prefer not to gamble even if the expected outcome is the same; the certain option has a greater expected utility. Eraker and Sox also tested whether framing the question in terms of losses and gains would influence attitude to risk and found that individuals tend to be risk averse if the situation was framed as a gain and as risk lover if it was framed as a loss, confirming observations by Kahneman and Tversky (1979).

Certainty, isolation and reflection effects contribute to intransitive and inconsistent preferences and result in value functions that exhibit different attitudes to risk. The prospect theory has conceptual appeal in describing how people make choices and how they process complex information. The theory presumes human beings to have constrained decision-making capabilities and 'bounded rationality'. However, the theory is complex with weighting functions on probabilities and outcomes, which are likely to differ between individuals.

2.2.2.3 Regret Theory

Regret theory was put forward by Loomes and Sugden in 1982 as an attempt to predict the violations of EUT ex-ante (Smith, 1996). The theory was also offered as an alternative to the Prospect theory on the basis that it is much simpler and has greater appeal to intuition compared to the prospect theory which has many ad-hoc and complex assumptions (Loomes and Sugden, 1982). Loomes and Sugden (1982) believe that in addition to explaining the systematic violations of EUT, their theory indicates that, "*such behaviour is not in any meaningful sense of the word, irrational.*"

The authors claim that the violations of EUT axioms could be explained in terms of decision regret. Loomes and Sugden (1987) explain that, *"the basic idea behind regret*

theory is that, when making decisions, individuals take into account not only the consequences they might experience as a result of the action chosen, but also how each consequence compared with what they would have experienced under the same state of the world had they chosen differently". The theory posits that the psychological experience of pleasure associated with the consequences of good A will depend not only on the nature of good A, but also on the nature of good B. If A has more desirable consequence, the individual will experience rejoicing and if vice versa, regret (Loomes and Sugden, 1982). These experiences happen after an individual reflects on their choices. The extent of experiences of rejoicing and regret depends on the choice-less utility function i.e. 'what is' and 'what might have been' and is independent on any other characteristics of the consequences. If what occurs is as pleasurable as what might have occurred, there is neither regret nor rejoicing and regret theory approximates the EUT.

Regret theory assumes that individual choice decisions are determined by the desire to maximize the net advantage of choosing A and rejecting B, in the event that state j occurs (O'Brien, 1990). Thus alternatives are valued simultaneously, which breaks the axiom of independence and provides a rationale for breaking the axiom of transitivity (Smith, 1996). Regret theory allows non-transitive pair-wise choices, as the relation of weak preferences is not necessarily transitive.

Noting that potential regret in health care is significant as decisions could mean literally life or death, Smith (1996) comments that conceptual and empirical interest of regret theory in health care is only tentative. In a study that used the TTO valuation technique to value outcomes from treatment of colon cancer states, i.e. cancer recurs (die) or does not recur (live), following surgery alone and surgery plus chemotherapy, Smith (1996) observed that regret was an important element in individual valuation and decision making.

The regret theory is criticized for not being explicit about how individuals form expectations or anticipation about regret as regret is an ex-post phenomenon yet the theory is about choice and the ex-ante capacity to anticipate feelings of regret and rejoicing (O'Brien, 1990; Keasey, 1984). The theory is based on non-observable functions in contrast

to the EUT, although its proponents note that it is possible in principle to infer from observations of individual choice, whether they are behaving according to regret theory (Loomes and Sugden, 1982). The theory is however simpler and does not impose the unnecessary restrictive notion of rationality.

2.2.2.4 Disappointment Theory

Disappointment theory is a variation of the regret theory, where the individual evaluates the actual outcome relative to expected outcome (O'Brien, 1990). Disappointment theory incorporates disappointment and elation and these emotions are postulated to follow verification or falsification of an expectation (Bell, 1985; Brandstatter and Kriz, 2001). The source of disappointment and rejoicing is the difference between the actual and expected outcome. Disappointment arises if expectations have not been met, and elation otherwise.

Disappointment theory examines utility gained form the same action in different states of the world as opposed to regret theory which examines utility from different actions in the same state of the world (Freemantle, 1996). Hence, unlike regret theory where final asset positions are evaluated irrespective of their probabilities, the source of (dis)utility in disappointment theory is evaluation of actual versus expected outcomes which are probability weighted (O'Brien, 1990). Freemantle (1996) notes that in making decisions for individual patients under uncertainty, health care professionals may be influenced by regret while their patients may be influenced by expected disappointment.

2.3 CONCLUSION

This chapter has reviewed the welfare theory together with theories of consumer choice under certainty that form the foundation of welfare economics. Also reviewed were theories of consumer choice under uncertainty, in recognition of existence of risk and certainty in health care decision making. Theories of consumer choice under risk and uncertainty are mostly formulated within the framework of expected utility. In attempts to understand the violations of the axioms of EUT, alternative theories have suggested different phenomenon

that may lead to consumers being 'irrational', intransitive and inconsistent, even though they would wish to act according to EUT axioms. These include the certainty, isolation and reflection effects of the prospect theory and the regret, disappointment, elation and rejoicing of the regret and disappointment theories. The prospect theory in particular shows that individuals display different attitudes to risk resulting in value functions that are concave over gains and convex over losses and are steeper for losses than for gains. Citing Gafni and Torrance (1984), O'Brien (1990) notes that risk attitudes in health choices can be split into quantity, gambling and time effect. The quantity effect is based on the concept of diminishing marginal utility and results in a concave value function, while the gambling effect assumes that individuals dislike gambles and time effect assumes that goods received in time t are valued more than in time t+1. While risk attitudes can be either risk aversion, neutral or seeking, it appears that individuals display attitudes that largely tend towards valuing certainty. In addition in cases of risk neutrality, there is little role for uncertainty. Therefore, it would appear that theories of consumer choice under certainty have a role in understanding consumer choices and are perhaps more useful in understanding consumer valuations of health and health care, from which their risk attitudes can be isolated. Hence consumer theories under certainty and uncertainty are complementary in aiding descriptions of consumer behaviour in making choices.

CHAPTER 3

CONCEPTUAL AND METHODOLOGICAL FRAMEWORK

3.0 INTRODUCTION

This chapter highlights conceptual and methodological issues in measurement and valuation of health related quality of life. It explores methodological issues surrounding measurement and valuation of health outcomes in general with special attention to disease specific outcomes. Identification of methodological gaps in measurement and valuation of health outcomes will contribute to setting and justifying the scope of the empirical work presented in chapters five to eight. The chapter aims to demonstrate the need and relevance for measurement and valuation of disease specific outcomes for use in economic evaluation in a developing country like Kenya.

Why Measure HRQL?

Measures of health outcomes have many uses that aid health care decision-making. These include: enabling description of the health of population (Dolan, 1997: Ebrahim, 1995: Patrick et al. 1993: Wilkin et al. 1992: Revicki et al. 1993: Kaplan et al. 1993: Guyatt et al. 1993); discriminating between patient groups and predicting possible outcomes (Wilkin et al. 1992: Bowling. 1997: Tolley et al. 1994); patient care; and, evaluating the effectiveness and efficiency of health care interventions (Guyatt et al. 1993: Wilkin et al. 1992). Measures of health outcomes are diverse, reflecting differing views in conceptualization of health and the changing concerns regarding the nature and assessment of impact of health care interventions (McDowell and Newell, 1996; Bowling, 2001 and Bowling, 1997). McDowell and Newell (1996) observe that "*the resolution of one type of health problem reveals a new layer of concerns,*" underscoring the fact that health indicators are continuously evolving together with their measurement to reflect prevailing concerns.

Several arguments support incorporating HRQL within the measurement of health outcome.

Increase in life expectancy has rendered traditional measures, such as infant mortality, imperfect indicators of the health of the surviving population (Patrick et al. 1993), as they ignore the status of the living (Bowling, 1997). Increased life expectancy has also been accompanied by a higher prevalence of disability in the population (McDowell and Newell, 1996; Patrick et al. 1993 and Wilkin et al. 1992), raising concerns of whether the extra years are spent in good or poor quality health (Bowling 2001). Quality of life as a measure of outcome re-directs attention towards consideration of impact of the condition and treatment on patient's emotional and physical functioning and lifestyle. It helps to answer questions of whether the treatment leads to a life worth living.

Incurable diseases have emerged as morbidity and mortality rates have fallen. The shift from acute life threatening conditions to chronic illnesses necessitates a different approach to the measurement of need and outcome, where the notion of severity becomes crucial (Wilkin et al. 1992). This is reflected by the increased emphasis in some countries on preventing ill health and disability, reducing health disparities between population groups and improving the quality and not just the quantity of life (Patrick et al. 1993; McDowell and Newell, 1996). The changing health problems and people's perception of their impact on their lives has called for new ways of measuring the outcomes of diverse health care interventions.

Purchasers of health care are expected to allocate scarce health care resources on the basis of evidence of cost effectiveness of health care interventions. To this end, health effects ought to be assessed in the broadest sense possible as many health care interventions are intended to improve general health and quality of life (Bowling, 2001).

Although the primary goal of medical treatment and care are to increase survival and add quality to the survival, many health care programs and interventions will have little impact on mortality (Bowling, 1997). For most interventions, the goal of treatment is palliative and

at times treatment only has impact on quality of life (Brown et al. 2000; Lee et al. 2001; Guyatt et al. 1999; Yee, 1997; Kerrigan et al. 2000; Douzdjian et al. 1998, Blumenschein and Johannesson, 1998; Bayoumi and Redelmeier, 1999 and Leung et al. 1999). Taking into account patients' preferences with regard to QOL improvements associated with a treatment, best captures these QOL effects.

The first two justifications for measuring HRQL may not apply today to developing countries like Kenya, which are still grappling with low life expectancies and high mortality rates[13] (MOH, 2002). However, both chronic and acute conditions prevail in these countries[14], making quality of life an issue of concern among the affected populations. For example, *S. Mansoni* with both acute and chronic stages is a disabling condition with long term quality of life implications, which have not been assessed to date. Furthermore, these countries are bereft of resources, which justifies appropriate health outcome assessments for aiding health care decision making.

Current interest by health care policy makers, professionals and patient groups is in measures of health that take into account both positive and negative effects of treatment or conditions and express them in terms of quality and quantity gains. QALY type measures have been developed in response to this need. They facilitate the combination of dissimilar gains in quality and quantity of life from health care interventions into a single measure and require both measurement and valuation of health outcomes. Within economic analyses such an approach can allow cross-program comparisons and inform on issues of technical and economic efficiency of health care programs. This information can usefully inform resource allocation decisions at national or international level.

[13] The infant mortality rate stands at 71 per 1000 births, maternal mortality rate at 590 per 100000 live births and the under five mortality rate at 112 per 1000 births (MOH, 2002). Life expectancy at birth in 1996 was 58 years (World Bank, 2003).

[14] For example, preventable and readily treatable diseases such as TB, typhoid, cholera, malaria and pneumonia as well as non-treatable diseases such as HIV/AIDS constitute the biggest burden on Kenya's Ministry of Health (MOH, 2003), with HIV/AIDS threatening to consume 50% of public health resources. Efforts to combat these health problems aim to increase longevity and quality of life of Kenyans (MOH, 2003).

There is a lack of distinction in usage of HRQL measurement and valuation instruments (Bowling, 2001), with some authors e.g. Guyatt et al. (1999) treating valuation instruments as measurement instruments. It is vital to keep the two steps of measurement and valuation separate (Drummond et al. 1997) for clarity about what is being measured and valued and also consistency with requirements by economists for an index-based type of measure.

The rest of this chapter is organized as follows. The next section looks at the conceptual and methodological issues in measurement of HRQL. In section 3.2, issues in valuation of health/disease states are addressed and section 3.3 presents the conclusion.

3.1 CONCEPTUAL AND METHODOLOGICAL ISSUES IN MEASUREMENT OF HRQL

Measurement is the activity or process of finding the size or amount of something by comparing it with a standard unit (Hornby, 1995). Measurement of health begins by defining the health concept. Different conceptions of health exist (Allen et al. 1997). However, the most widely invoked concept of health is WHO's (1993) as a "complete state of mental, physical and social well-being and not merely the absence of disease and infirmity", although Kaplan and Anderson (1988) observe that WHO neglected to provide operational definitions. Depending on the definition or concept of health adopted, health is operationalized and measured by describing levels of functioning on various dimensions or domains.

Lately, there has been increased emphasis on measurement of health status, quality of life and health related quality of life. However, the distinctions between the three concepts are not clear (Shumaker and Naughton, 1995), and they are often used interchangeably in the literature, occasioning a great deal of confusion. The three concepts are briefly described below together with the distinction between disease and health as the interest of this thesis focuses on disease. This clarifies what each of the concept entails and shows how intricately intertwined they are.

3.1.1 QOL, Health Status and HRQL

Encompassed in the term HRQL are two broad concepts; health and quality of life. Shumaker and Naughton (1995) note that while these concepts are multidimensional, the key dimensions remain in dispute and controversy persists in the HRQL field.

Quality Of Life (QOL): The concept of quality of life is complex, broad and involves highly subjective value judgements. There is no agreement (Bowling 1997; Hunt 1997) about the meaning of the concept and it has been described using terms such as health status, physical functioning, perceived health status, subjective health, health perceptions, symptoms, need, satisfaction, individual cognition, functional disability, psychiatric disturbance, well-being and often several of these at the same time. Other key descriptors of quality of life include social well being, personal esteem and satisfaction, emotional and economic status, happiness and overall satisfaction with life. Though amorphous and vague in nature the term is multidimensional and theoretically incorporates all aspects of an individual's life (Bowling, 2001). Hence the concept is considered to be broader than personal health status. Although this term may be referred to occasionally in the text, it will not be investigated in this study in its broader sense.

Health Status (HS): Most of the meanings ascribed to health status draw from WHO's (1993) concept of health (Wilkin 1992). Key terms used in describing the concept include well-being, functional status, maintenance of strong social support system and integration in the community (Bowling, 1997: Patrick et al. 1993: Wilkin et al. 1992: Guyatt et al. 1993: McDowell et al. 1996). It also includes levels of physical fitness and physical health, achieving functional excellence, ability to cope with stressful situations and psychological well being. Other terms include developing full human potential, maximization of ones quality of life as well as absence of disease, impairments, and handicap (Bowling, 1997: Patrick et al. 1993: Wilkin et al. 1992: Guyatt et al. 1993: McDowell et al. 1996). Hence good health status can be seen as part of improving the quality of life.

Health Related Quality of Life (HRQL): Patrick et al. (1993) proposed a conceptualisation of health related quality of life as "*value assigned to a duration of life as modified by the impairments, functional states, perceptions and social opportunities that are influenced by disease, injury, treatment or policy*". This draws attention to the relationship between quality and quantity of life, the multi-dimensional nature of quality of life and emphasises the need for a measure that is sensitive to variation in disease, treatment and policy. Hence, conceptualisation of health related quality of life appears to imply that only those aspects of quality of life affected by disease, injury, treatment or policy should count in assessing the concept. It combines both the biomedical and the quality of life concepts of health. Hence, in assessing the impact of disease, injury, treatment or policy on health related quality of life, one can assess in terms of impairment arising from the condition, as well as impact of the disease on other aspects of quality of life directly affected by the disease. Shumaker and Naughton (1995) observe that Patrick et al's definition provided limited information on the dimensions that constitute HRQL and propose a definition that makes explicit the key dimensions of HRQL as referring to "*...people's subjective evaluations of the influences of their current health status, health care, and health promoting activities on their ability to achieve and maintain a level of overall functioning that allows them to pursue valued life goals and that is reflected in their general well-being. The domains of functioning that are critical to HRQL, include: social, physical and cognitive functioning; mobility and self-care; and emotional well being (p.7).*"

While the three terms (QOL, HS, HRQL) may be used interchangeably, the usage of the terms and investigation for the purpose of this study will draw from the definition of health-related quality of life by Patrick et al. (1993) and further clarified by Shumaker and Naughton (1995) and Bowling (2001, p. 6) as mirroring WHO's definition of health.

Disease versus Health: In assessing health outcomes, the constituents of the instrument reflect the conception of health adopted. The conceptualization and measurements of outcomes is a controversial area largely due to lack of exact definitions of the concepts.

Most indicators of health such as mortality and morbidity rates reflect a 'disease' model (Bowling, 1997), concerned with pathological abnormality as indicated by signs and symptoms and detectable by medical science. Disease is illness of the body, or of the mind caused by infection or internal disorder and given a specific name.

Focusing on disease tends to reinforce the use of medical approach to interventions where the objective is to eliminate the disease. This is a somewhat narrow view considering that apart from eliminating disease and its accompanying consequences on the persons life, most health care interventions will have an impact on other important aspects of patients' life, but which this model largely ignores. As Wilkin et al. (1992) observe, the medical model considers health to be the absence of disease and measures health in terms of morbidity, mortality, incidence and prevalence of disease as well as death rates. This model makes little appeal to the consequences of disease (Wilkin et al. 1992, p.11) and it has no regard to how the patient feels. Incorporating concerns and feelings of the patients shifts focus from 'disease' to consequences of disease and into the realm of the broader concept of health.

The term health can be seen as a combination of complex phenomena that goes beyond the traditional medical model (disease and infirmity) to a model that evaluates the consequences of the disease for the person in totality. Different conceptions of health exist (Allen et al. 1997)[15] and the concept is subject to different cultural interpretations (Bowling, 2001) as shaped by beliefs and behaviours of study participants, although researchers have largely influenced the domain content of the concept (Bowden, 2001). Fox-Rushby and Parker (1995) show that different generic HRQL instruments use different definitions of health and therefore have different dimensions representing HRQL. For example the WHOQOL includes the dimension of spirituality that is not found in other instrument. Bowden (2001) and Fox-Rushby et al. (2001) highlight differences in conceptions of HRQL represented in existing HRQL instruments developed mainly in

[15] Allen et al. (1997) offer six different ways health is conceptualized. Their discourse portrays health as ranging from one extreme of objective truth, through health as an English word like any other, health as absence of bodily malfunction, local equivalence with connotations of health the English word, local ways of seeing the world with ways implied by English word health to health as a word that relates only to local experiences (extreme cultural relativism).

Western Europe and North America and the KENQOL, developed amongst the Wakamba in Kenya.

3.1.2 Measurement instruments

Health can be measured using one or a battery of the many health measurement instruments that exist. Depending on the scope of the instrument, a distinction is made between disease-specific and broad-spectrum generic health measures.

Generic measures are useful due to their broad nature and ability to inform a wider variety of resource allocation decisions as they allow comparability across different diseases and populations (Patrick and Deyo, 1989; Brazier et al. 1999. However, they are criticized for being irrelevant (Brazier et al. 1999 and Brazier and Dixon, 1995) when applied to specific diseases as they may fail to capture all aspects of a condition and are often insensitive to small but clinically important changes in patient's health status (Gold et al. 1996). Such changes may havea social impact that matters to patient. Disease-specific measures become useful when disease related attributes need to be assessed and greater sensitivity is desired, since they focus on the relevant aspects of different conditions (Bowling, 2001; McDowell and Newell, 1996; Brazier et al.1999; Cairns, 1996; Brazier and Deverill, 1999). They however have limitations in comparisons across diseases curtailing their usefulness in economic evaluations (Cairns, 1996; Brazier et al. 1999; Cairns, 1996; Chancellor et al. 1997 and Brazier and Deverill, 1999), when the disease states they measure are not valued. A common view in the literature is that both generic and disease specific instruments should be used alongside each other where logistics allow, as they seem to complement rather than substitute each other.

Health-related quality of life instruments can be index (utility) or profile producing and generic or disease-specific. The choice of an instrument depends on intended use as well as other factors such as psychometric properties and practicality in use. Preference-based index producing HRQL instruments are most suitable for economic analysis, which is the discipline of this study. Three most commonly used and well established generic index-

based health status classification systems (HSCS) were identified in the literature[16]. They are EQ-5D, (EuroQOL group, 1990: Brooks and EuroQOL Group, 1996), HUI 1 to 3 (Health Utilities Index) (Torrance et al. 1996: Furlong et al. 1998, Torrance, 1996[17]) and QWB (Quality of Well-Being) (Kaplan and Bush, 1976: Kaplan et al. 1984: Read et al. 1987: Kaplan and Anderson, 1988). The SF-6D is a recently developed method for deriving a single index value from the SF-36 classification system (Brazier et al. 1998; Brazier et al. 1999b). These instruments provide an avenue for obtaining health state scenarios from the population which can be valued either indirectly using the scoring formula that is part of the system or directly by obtaining preferences from a sample of the population using the valuation methods described in section 3.2. Other than the patient utility measurement scales (PUMS) used in arthritis drug trial by Bombardier et al. (1986), there appears to be no other disease-specific measure that also produces utilities.

Generic and disease specific health state measurement instruments whether profile or index producing, first give descriptive information about health/disease states. This information tells us the "amount of various health/disease states" in a given population. This is the step that constitutes health/disease status measurement and essentially, any of the existing health/disease measurement instruments can do this. Different health/disease state instruments produce diverse descriptions of health/disease. These may be seen as disparate 'physical units' of health/disease whose values are not ascertained. At this point, however, we can not tell what value or worth any given health/disease state has to the individual or society, as elicitation of values for different health/disease states is required to ascertain their worth.

To date, none of these health status classification systems have been tested and or applied in Kenya to measure and describe population health and various patient groups. Their use in such settings would require testing them for equivalence as espoused in Herdman et al. (1998) to decide whether entirely new instruments are needed or whether the existing ones

[16] Other less commonly used index-based health status classification systems include Rosser Disability/Distress index, 15D , Quality of life and Health Questionnaire and Years of Healthy Life Measure (Gold et al. 1996. P. 124-128)

[17] Development of HUI mark 1 to 3 has been undertaken by researchers in Center for Health Policy Analysis (CHEPA), McMaster University, Canada.

can be adapted for the new setting. However, Fox-Rushby and Parker (1995), Bowden (2001) and Bowden and Fox-Rushby (2003) suggest that existing measurement tools could not be used, which necessitates development of new tools that account for local perceptions of HRQL. In the next two sections approaches to developing new HRQL measures and issues in cross-cultural adaptation of existing measures are considered. Issues of equivalence in cross-cultural adaptation of instruments are also relevant to development of new tools and are only considered in section 3.1.4 to avoid repetition.

3.1.3 Approaches to Developing HRQL Measurement Instruments

Development of HRQL measurement questionnaires consists of several steps namely: item generation and ascertaining their content validity; item reduction and development of scales; and finally testing for validity, reliability, responsiveness, sensitivity and acceptability (Cano, 2001). Streiner and Norman (1995) offer a detailed description of each of these steps. How elaborate each of these steps gets depends on time, expertise and resource availability (Guyatt, 1995) and perhaps the nature of instrument being developed (Fayers and Hand, 2002).

Item generation and ascertaining of content validity is an important step as it focuses on ensuring that the instrument is able to measure what is intended accurately. Items are normally generated from literature reviews (Bullinger et al. 1993; Guyatt, 1995; Wu et al. 1995), patients (Arpinelli, et al. 1995; Avis and Smith, 1995; Girman et al. 1995; Lara-Munoz et al. 1995; Marquis et al. 1995a; Marquis et al. 1995b), specialists or clinical observation (Marquis et al. 1995c; Streiner and Norman, 1995) and well people from the community (Skevington et al. 1995). Other sources of items include theory, research and expert opinion (Streiner and Norman, 1995). The methods followed in generating the items include focus group discussions, which help in suggesting general themes and discussing whether items are relevant, clear and unambiguous as well as key informant interviews. Having generated a pool of items, panels of experts and patients ascertain content relevance by checking whether all concepts important for the patient are measured (Marquis et al.

1995a), and suggesting additional items to fill any important areas that may have been missed.

Item reduction and development of scales is achieved through patient perceived importance via rating of importance of items (Girman et al. 1995), descriptive statistics such as frequencies (Marquis et al. 1995a) as well as principal component analysis, multi-trait analysis and step-wise discriminant analysis. Items are dropped according to pre-set criteria (Marquis et al. 1995a and 1995b). Item reduction and development of scales is accomplished through administering the questionnaire to a sample of target group. The final step involves testing for psychometric soundness of the instrument using a variety of analytical tools (Streiner and Norman, 1995; Carmines and Zeller, 1979; Brooks, 1995; McDowell and Newell, 1996).

Although the above steps form the basic structure in instrument development, there has been criticism regarding the viewpoint taken in conceptualizing health/disease embodied in the instruments. These criticisms relate to the ways item and content validation of the instruments has been undertaken in the past (Bowden, 2001). Bowden (2001) and Fox-Rushby and Parker (1995) note that researchers have dominated development of instruments and that many of the concepts held by the lay population are not represented in HRQL measures. A possible reason for this situation is that instrument developers have not sought to establish the local conceptualization of health/disease before devising items and developing measures (Fox-Rushby, 2002). This has implications for the future assessments of conceptual equivalence of instruments if transferred to other cultures, as it lays the foundation and basis for comparisons.

Items in disease specific instruments often consist of symptoms and their impact on patient's functioning. They are often chosen, *"on the basis that they constitute an important aspect of the concept"* and are therefore intrinsic to the definition of that concept (Fayers and Hand, 2002). Such items are referred to as causal variables (i.e. experience of a symptom may cause low HRQL) and are less homogeneous when they constitute an instrument than those found in psychometric tests. Because they are part of defining the

concept, psychometric tools for item reduction such as correlations should not be applied as a basis for retaining or dropping the items. This arises from the fact that items that cluster together in a clinically sensible manner are not necessarily highly correlated and therefore the correlations should not be expected to be high (Fayers and Hand, 2002). Fayers and Hand (2002) suggest basing item reduction on interviews with the patients to rate the importance and impact of these types of variables. The sense in this is that any one of the causal variables i.e. symptoms and side effects can lead to a poor HRQL (Fayers and Hand, 2002) and would therefore be equally important in the instrument.

Fayers and Hand (2002) suggest that in developing HRQL instrument where causal variables are present, there is need to combine both the psychometric and clinimetric approach to yield suitable measuring instruments. They define the clinimetric approach as, *"based on a deliberate choice of what variables to include and, in the absence of an underlying model, a deliberate choice of how these variables should be combined."* The clinimetric approach is distinguished from the psychometric approach in that in the latter all items reflect a single latent variable and are homogenous while in the former, items represent several distinct latent variables (Fayers and Hand, 2002).

3.1.4 Cross-cultural issues and equivalence in use of instruments

While many HRQL instruments exist, only a few have been developed with an international focus from the outset (Herdman et al. 1997). This has necessitated either development of new instruments as described in section 3.1.3 or use of instruments previously developed in another language and setting. Development of a new HRQL instrument is a time and resource consuming process and is usually undertaken when a suitable instrument for use in a new setting is lacking. Cross-cultural transfer of an already existing instrument raises issues of translation and cross-cultural adaptation of previously developed instruments for application in new settings due to language and cultural differences. Cross-cultural adaptation requires the translation of the HRQL measure and its adaptation with regard to idiom, cultural context and lifestyle so as to measure similar phenomena in different cultures (Guillemin et al. 1993). Guillemin et al. (1993), Herdman et al. (1997 and 1998)

and Bullinger et al. (1993) have provided criteria for determining the suitability of instruments for translation as well as guidelines for cross-cultural adaptation of HRQL instruments. These include translations and back-translations, a committee review, pre-testing and weighting of scores. To investigate meaning and quality of the translated instrument, Herdman et al. (1997) and (1998) suggest fulfilment of six types of equivalence. These include conceptual, item, semantic, operational, measurement and functional equivalence and are briefly discussed [18].

Equivalence of Instruments: International comparability of HRQL outcomes across and within countries, health care interventions, clinical trials and disease conditions requires comparison of like with like. This implies that the concept being measured, the HRQL, ought to have the same meaning across cultures and nations (Bullinger et al. 1993). As Herdman et al. (1997) argue this has not always been the case, in that there lacks standardization in usage of terms and methodologies (Guillemin et al. 1993: Bullinger et al. 1993) for cross-cultural adaptation and in some cases development of new instruments[19]. However, Herdman et al. (1998) offer a model of assessing equivalence in HRQL instruments both in developing new measures and in cross-cultural adaptation of existing HRQL instruments. The model consists of six types of equivalence and begins with assessment of conceptual equivalence.

Conceptual equivalence involves exploring ways in which health and quality of life are conceptualized as well as values placed on health and quality of life. Conceptual equivalence can be assessed by examining perceptions of health, illness, disease, quality of life either by consulting experts or the general population in terms of beliefs and behaviors regarding health and quality of life. Once domains are identified, people can be asked to prioritize them in order of importance (Herdman et al. 1998). This is considered the first and perhaps the most important step in questionnaire development and adaptation. It deals with identifying and describing the concept to be measured.

[18] Refer to Herdman et al. 1997 and 1998 for more details.

[19] HRQL instruments have not always been developed with an international focus. Most are developed in English (UK and North America (Canadian English)) and often it is not clear how items of the questionnaire were chosen (Bowden, 2001; Rosenbaum and Saigal, 1996).

Item and semantic equivalence should be explored once conceptual equivalence has been established. They are concerned with suitability of items and ease of translation of items in order to transfer the same meaning. Item equivalence can be established by consulting members of the population, examining patterns of lifestyles and habits or asking experts. Semantic equivalence can be established through understanding of key words and the descriptions behind the key words by developers of instruments (although they don't always exist). In developing a new instrument it is important to provide the descriptions of key words and phrases that comprise the instrument.

Operational equivalence concerns the mode of interview format, ordering of items, mode of answering and time frame. This is explored by finding out how the questionnaire has been administered. In case of a new instrument this may involve assessing the level of education of respondents and testing the proposed method with a sample of the proposed population and their familiarity with the intended mode of administration. It is also important to assess whether there are differences in results between modes of questionnaire administration (Herdman et al. 1998). The fifth type of equivalence concerns the psychometric properties of the instruments such as validity, reliability, sensitivity and responsiveness. Finally, examining the level or degree to which other equivalences are achieved assesses functional equivalence. At each level of assessment, the decision whether or not it is appropriate to adapt an instrument is made while attempting to get "as close as possible to exact equivalence" (Herdman et al. 1998).

The Herdman et al. (1998) guidelines were operationalized for the first time in Bowden and Fox-Rushby (2003), where the processes used in translating and adapting nine generic HRQL instruments in Africa, Asia, Eastern Europe, the Middle East and South America was systematically and critically reviewed. With the exception of the WHOQOL, Bowden and Fox-Rushby (2003) reported poor levels of assessment of equivalence for other measures considered and call for change in research practice and translation guidelines to facilitate more effective and less biased assessment of equivalence of HRQL measures across countries.

The guidelines, referred to as the universalist approach to health related quality of life are reflected partly in the on-going HRQL work amongst the Akamba of Kenya towards development of a generic HRQL instrument, the KENQOL (Fox-Rushby, 1994, 1995: KENQOL Group, 1996). The KENQOL group used semi-structured interviews and focus group discussions as well as protracted participant observation fieldwork (Kirstin, 1999; Fox-Rushby et al. 1995; Amuyunzu et al. 1995). This aimed at providing information of general perceptions of health and well-being in the community, personal accounts of health, illness and well-being in the individual and family and general descriptions of the local area (Fox-Rushby et al. 1995) which helped in addressing themes previously ignored in previous HRQL literature. Bowden et al. (2002) further report on qualitative methods for pre-testing and piloting the survey questions to establish the validity and reliability of the instrument before subjecting it to item reduction and quantitative psychometric testing for validity and reliability (Fox-Rushby et al. 2001). The methods adopted in the development of the KENQOL instrument enhance its future use in other settings because they provide a basis for judging the conceptual and other forms of equivalence of future translations (Bowden et al. 2002). Although this approach requires enormous amounts of time, different expertise and financial resources, it is most suitable for developing generic HRQL measure, because it is grounded on local conceptualization of health, well being, illness and disease and therefore helps in defining what is being measured.

At present, no HRQL instrument specifically developed for Kenya exists and although the KENQOL is currently being developed (Fox-Rushby et al. 2001; Bowden et al. 2002; Nyandieka et al. 2002) it is not ready for use. Therefore, at the moment application of a HRQL instrument in Kenya to classify people into various possible health-states requires either cross-cultural adaptation of existing instruments or development of instruments specific to Kenya.

3.2 ISSUES IN VALUATION OF HRQL

Valuation is the action or an instance of estimating or judging the quality or worth of something or somebody (Hornby, 1995). In this case valuation refers to judging the quality of or worth of a health state. The relevance of valuation to HRQL lies in that, changes in health states can be counted (through measurement) and their worth determined (through valuation) as an index. The benefit in obtaining values for health state descriptions is that issues of technical and economic efficiency can be addressed because outcomes from different treatments and interventions are expressed in the same units. To the extent that similar and different health states are assigned different values using same numeraire, (be it quality weight (QALY, HYE), disability weight (DALY,) or monetary value (WTP)), valuation offers an opportunity for ascertaining the value of the outcomes and thereby a way of bringing together disparate 'physical units of health states' into a common measure that can be used for economic evaluation of intervention programs.

Valuation of health is central to the construction of the QALY type outcomes, a process that makes several assumptions and value judgements. For example, the process of valuation of health states involves issues such as describing health states in various domains, deciding on the type of valuation techniques and valuation questions to be used, choosing raters and deciding on the range of health states to be valued. To obtain preferences that can be used in health care decision making, it is assumed that: individuals are rational and are able to express their preferences; their preferences remain constant overtime and preferences for health states are independent of the duration of state; their attitudes to risk are neutral and constant. It is also assumed that preferences are cardinal and therefore individual preferences can be aggregated and interpersonal preference comparisons made for 'collective priority-setting' (Williams, 1996). Each of these issues are matters of extensive debate in the literature (Torrance, 1986; Brooks, 1995; Drummond et al. 1997; Gold et al. 1996; Williams, 1996; Carr-Hill, 1989 &1992; Broome, 1993; Mulkay et al. 1987), with some questioning and others defending the usefulness of valuation for guiding decisions in health care.

Different health state valuation techniques produce different values for similar and different states (Dolan et al. 1996a; Drummond et al. 1997; Read et al. 1984; Stigglebout et al. 1994; Rutten-van Molken et al. 1995). This is not surprising given that each technique frames the valuation questions differently and use different calibrators i.e. time, probability, money and numerals. Hence methods are choice-based or choice-less and either incorporate certainty or uncertainty. Also, different techniques have different theoretical explanations. Choice of which valuation technique to use has been debated (Torrance, 1986; Drummond et al. 1997; Gold et al. 1997; Brazier et al. 1999b) based on the criteria of performance in terms of validity, reliability and practicality. While no one method has unanimously been recommended, preference for choice-based methods has been voiced (Brazier et al. 1999b) although choice-less methods like VAS are also considered on the basis of being less costly.

Arguments abound regarding whose values should be used, amongst health professionals, general population and patients (Torrance, 1986). Gold et al. (1996) recommend that the choice of raters should be guided by the purpose and perspective of the analysis, such that where patient clinical decision making is at hand patients values should be used and where public resource allocation and planning decisions are relevant, values from the general population should be used. Mulkay et al. (1987) and Carr-Hill (1989) takes issue with whose values are used, noting that "the patients' responses are to questions framed and presented by the health economist and not as patients would themselves have formulated them". Mulkay et al. (1987) and Carr-Hill (1992 & 1989) therefore question the value judgements that go into deciding the raters as well as the content of health states that are valued. In the absence of suggestion of a better option by Mulkay et al. (1987) and Carr-Hill (1992 & 1989), Williams (1996) recommends being explicit about what the values are and who they are from.

The assumptions mentioned above are critical in obtaining values for health states. These assumptions are restrictive and research has shown their violation (see chapter 2), which has led to formulation of alternative ways of explaining the violations and understanding the process of preference formulation. Critics of use of health state values such as Mulkay

et al. (1987), Carr-Hill (1992 & 1989) and Broome (1993) question whether, given these violations, "valuations derived in an experimental situation can be applied in practice" and whether values for health states have any meaning. Their use is further questioned on grounds that studies eliciting health state values often have low response rates and exclusion of those who refuse to answer such questions as the tasks demand because they are apparently labeled 'irrational' (Carr-Hill, 1992). This touches not only on the generalizability of health state values but also raises questions whose views and values judgements are being represented. Related to this, is the issue of aggregation of values across individuals and interpersonal comparisons of utility.

Due to variation of preferences for health states by individuals and over time, gains from health interventions will vary according to tastes and preferences (Carr-Hill, 1989). Therefore, taking averages of preferences as is done in health state valuation misrepresents the views of some sufferers and denies them the treatments commensurate with their preferences. It also neglects other externalities such as the effect medical interventions have on others and fails to recognize that a year of healthy life can vary between individuals at different life stages e.g. the old and young, gender and family circumstances. While these concerns are legitimate, Williams (1996) notes that for collective decision-making, some form of aggregation is needed and advises explicitness in aggregation rules so that the ethical implications can be clarified. To the extent that preferences can be shown to be on a cardinal scale, preference aggregation and interpersonal comparisons can be undertaken (Broome, 1993) bearing in mind the difficulties involved.

By "putting different QALYs together we are making interpersonal comparisons" and assuming that QALYs represent that same value to different people (Broome, 1993). This implausible assumption has been a basis of criticism of health state valuation on the grounds that it ignores distribution of health gains. However, although the pursuit of efficiency in health care implies maximization of benefits, Williams (1996) argues that a QALY has equal social value regardless as to who gains it and that distributional issues can be built into QALY construction through appropriate weighting.

The field of health state valuation has seen considerable growth over the last three decades (Drummond et al. 1997). Williams (1996), responding to much of the criticisms in the field cautions against rejection of the QALY approach simply because it falls short of perfection, until there are better alternatives available! Use of health state values to inform health care decision making ought to begin with ascertaining that the values are valid and reliable for the intended use and that the available tools for eliciting values are practical in settings they are used in.

The next section presents the commonly used non-monetary valuation techniques. This is followed by a review of the use and performance of each technique in high-income countries, an examination of the factors causing variation in values and cross cultural issues in valuation.

3.2.1 Valuation Techniques

Five non-monetary preference (utilities and values) elicitation techniques were identified in the literature. They are time trade-off (TTO), standard gamble (SG), rating scales (RS), magnitude estimation (ME) and person trade-off (PTO) previously referred to as equivalence of numbers. There is no consensus regarding the best method for eliciting preferences for health states (Drummond et al. 1997) for use in resource allocation. Preferences for health states can either be utilities or values (Drummond et al. 1997). The distinction between utilities and values depends on question framing. Question framing concerns whether there is certainty or uncertainty, where uncertainty captures the respondent's risk attitude. Response method concerns whether the respondent is asked to make a choice or not. Choice based methods are rooted in economics and decision sciences, while scaling is rooted in psychology and psychometrics. Table 3.1 taken from Drummond et al. (1997) clarifies these distinctions and the valuation techniques that fall under different categories. While all the methods can be used to elicit preferences, those that do not incorporate uncertainty produce values. The standard gamble incorporates uncertainty and therefore produces utilities of the von Neumann-Morgenstern type.

Table 3.1: Methods for Measuring Preferences

Response Method	Question framing	
	Certainty (Values)	Uncertainty (Utilities)
Scaling	A ♦ Rating scales ♦ Category scaling ♦ Visual analog scale ♦ Ratio scale	B
Choice	C ♦ Time trade off ♦ Paired comparison ♦ Equivalence or Person trade off	D ♦ Standard gamble

Source: Drummond et al. 1997. pp 146

3.2.1.1 Standard Gamble (SG)

In SG, subjects reveal their indifference point between two alternative states, one risky and the other certain but lasting the same duration. The risky alternative has a specified probability of the more preferred health-state and a complementary probability of the less preferred health-state. To obtain utilities, the procedure usually begins with presenting a written description of the health-state (a scenario) to the individual whose opinion is being sought (rater). After reading the scenario, the rater is asked to imagine a hypothetical situation in which he or she is confronted with a choice. The options available are to continue living in the state of health described in the scenario, or to take a gamble. The gamble, which might be expressed as taking a medication or taking an operation, has two possible outcomes. The best outcome, is usually the restoration of perfect health and the worst outcome is immediate death. The last step in the SG is to systematically vary (in a Ping-Pong version, Furlong et al.1990) the probability (p) of attaining the best outcome of the gamble until a point is reached where the rater is indifferent between continued life in state being valued and taking the gamble. The value (p) is referred to as the rater's indifference probability. Utilities can be obtained for chronic health states preferred to death, states worse than death and temporary health states (Torrance, 1986) as illustrated in figures 3(a) to (c).

Until recently SG has been considered to be the gold standard in the valuation of health states (Torrance, 1987, Drummond et al. 1997, Torrance et al. 1996) due to its theoretical foundation in the von Neumann-Morgenstern expected utility theory (VN-M EUT). The standard gamble is one of the techniques preferred by economists because it involves choice (thereby invoking the concept of opportunity cost) and is framed in terms of risk (Bleichrodt, 2002), an attribute that characterizes most decisions in health care. The technique also has the strongest theoretical foundation (Brazier et al. 1999b). In addition, the SG has been found to be practical for use amongst a variety of populations and has reliability and validity. Some disadvantages of the SG have included the fact that: the procedure of utility elicitation is complex (Brazier et al. 1999b); it is time consuming to administer; that people have difficulties understanding the concept of probability; and that utilities can be influenced by the way questions are framed (Stavros, 2001).

The superiority of theoretical positioning of the SG amongst other valuation techniques has been questioned as the axioms of EUT are often violated when people make choices in practical situations (Fishburn, 1989; Smith, 1996; Wu, 1996; Cohen, 1996; Bernard, 1986; Llewellyn-Thomas et al. 1982). Section 2.2.2.1 presents the criticisms of the VN-M EUT arising from violation of the assumptions of independence and continuity, rationality, transitivity, consistency and consumer sovereignty. Alternative theories such as regret theory (Loomes and Sugden, 1982, 1987; Loomes et al. 1992; Sugden, 1989), prospect theory (Khanmann and Tversky, 1979) and disappointment theory (Bell, 1985) have been put forward to explain these violations.

Current literature has questioned the SG beyond its theoretical superiority by examining its usefulness in representing people's preferences for health for use in health care decision making. For example, Bleichrodt (2002) notes that under expected utility theory, the only reason SG and TTO utilities would differ is utility curvature, while empirical evidence abounds that EUT does not describe individual preferences well due to probability weighting, loss aversion and scale compatibility biases. Analyzing the effect of each of the biases, Bleichrodt (2002) finds that SG utilities will generally overestimate utility of a health state, while these effects lead to under- and overestimation in TTO depending on the

size of different biases. He therefore argues that elicitation of utilities should be based on the best descriptive technique available, i.e. where the joint impact of probability weighting, loss aversion and scale compatibility is minimized. The current (Dolan, 1996; Spencer, 2004) view is that taking these biases into consideration, the TTO better describes people's preferences for health than SG because, as Bleichrodt and Johannesson (1997) suggest, the downward bias in the TTO caused by utility curvature approximately offsets the upward bias caused by loss aversion and scale compatibility.

Recently, there has been a debate on the relevance of experimental economics suggesting the need for research and deeper understanding of factors behind preference reversals and the context within which experimental economics is carried out (Loomes, 1999: Starmer, 1999: Binmore, 1999: Loewenstein, 1999). In a recent paper, Baker and Robinson (2004) presented qualitative information surrounding construction of preferences using the SG and show that respondents incorporate a wide range of factors into their SG responses that include family circumstances, living situations and religious beliefs. They argue that these considerations may offer explanations of the inconsistencies found in preference elicitation studies and violations of the assumptions inherent in the theoretical basis of the techniques, such as utility independence and constant proportional trade-off. Robinson et al (2001) sought qualitative data that might provide insights into the thinking behind SG and VAS responses observed in their study and argue that reference point effects (as posited by Khanemann and Tversky's (1979) prospect theory), switch from choice-less (VAS) to choice-based (SG) method and differential weighting of individual attributes in a state are possible explanations of preference reversals. The growing evidence challenging the theoretical basis of the SG calls for different questioning to inform choice amongst valuation techniques such as evaluating the conceptual basis of the instruments in different cultures, in addition to practicality, ease of use and appropriateness.

Figure 3a: Standard Gamble for a chronic health state preferred to death

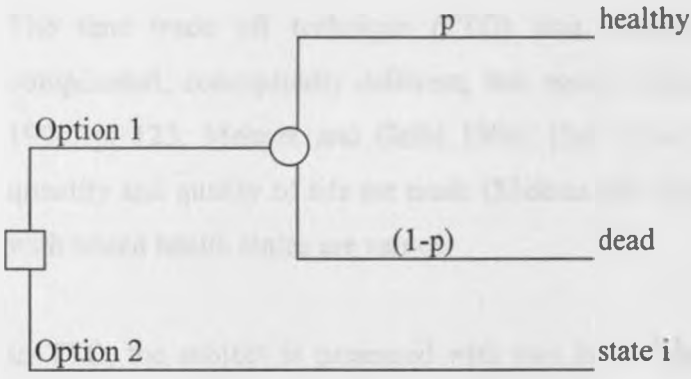


Figure 3b: Standard Gamble for a chronic health state considered worse than death

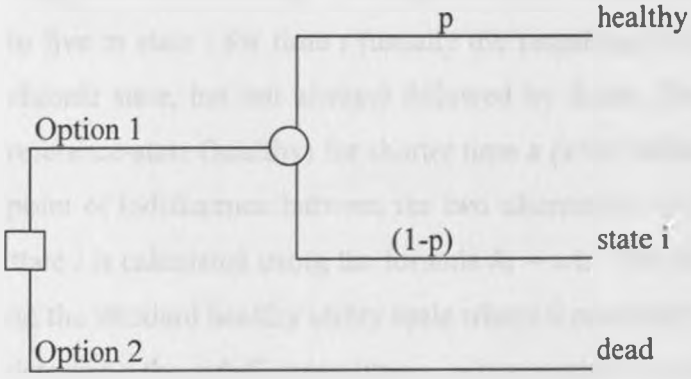
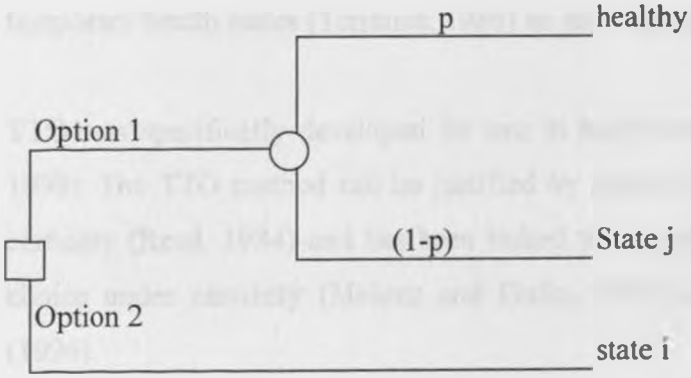


Figure 3c: Standard Gamble for a temporary health states i



3.2.1.2 Time Trade Off (TTO)

The time trade off technique (TTO) was developed by Torrance (1972) as a less complicated, conceptually different, but equally sound alternative to SG (Krabbe et al. 1996, p. 123, Mehrez and Gafni 1990, Unic et al. 1998). In TTO trade-offs between quantity and quality of life are made (Mehrez and Gafni, 1990) using time as the currency with which health states are valued.

In TTO, the subject is presented with two health-states, one of which is the state to be valued (state *i*) and the other is the reference or ideal state (usually healthy or best imaginable health state). The subject is then offered two alternatives. The first alternative is to live in state *i* for time *t* (usually the remaining life expectancy of the individual in the chronic state, but not always) followed by death. The second alternative is to live in the reference-state (healthy) for shorter time x $(x<t)$, followed by death. By varying time x, the point of indifference between the two alternatives is reached and the preference value for state *i* is calculated using the formula $h_i = x/t$. The calculated value can range from 0 to 1 on the standard healthy utility scale where 0 represents death and 1 represents 'healthy'. To determine the indifference time x, a converging Ping-Pong strategy is used, to negate the anchoring bias, which can otherwise confound the results (Mohide et al. 1988). Values can be obtained for chronic health states preferred to death, states worse than death and temporary health states (Torrance, 1986) as shown in figures 3(d) to (f).

TTO was specifically developed for use in health care (Torrance et al. 1972: Unic et al 1998). The TTO method can be justified by axioms of utility theory under conditions of certainty (Read, 1984) and has been linked to the indifference curve theory of consumer choice under certainty (Mehrez and Gafni, 1990) as illustrated by Buckingham et al. (1996).

Figure 3d: TTO for a chronic health state preferred to death

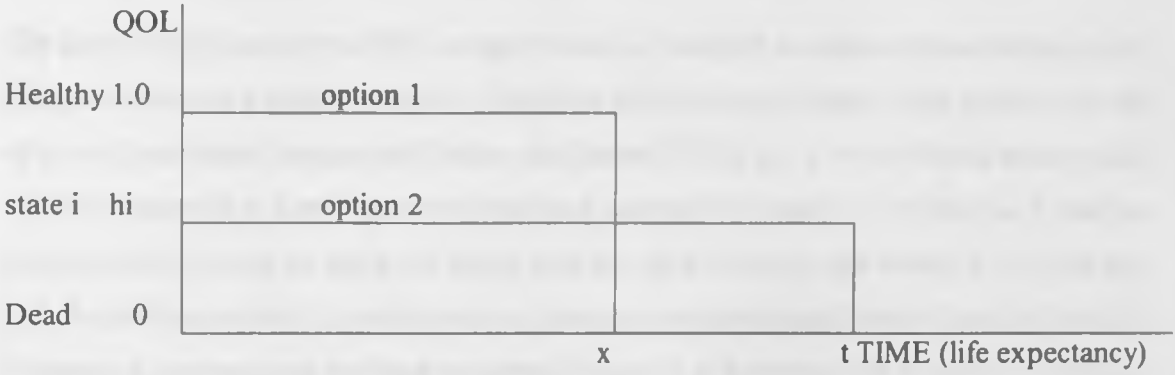


Figure 3e: TTO for a temporary health state

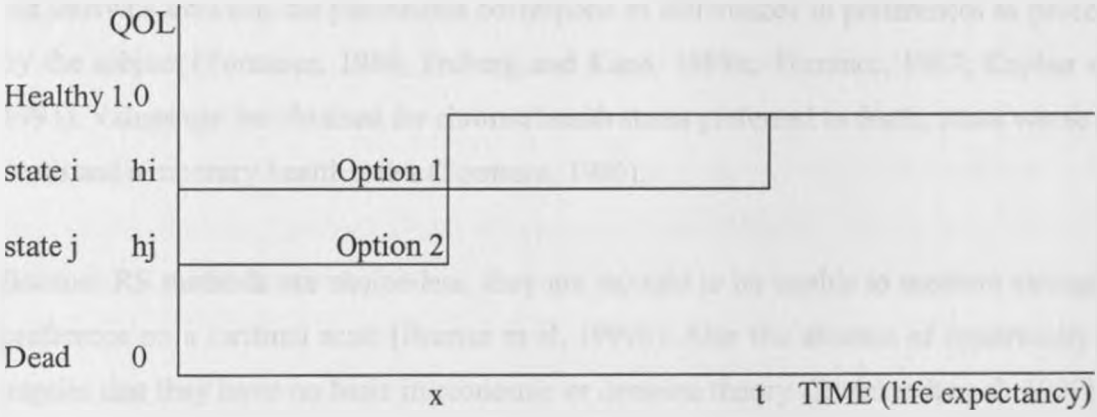
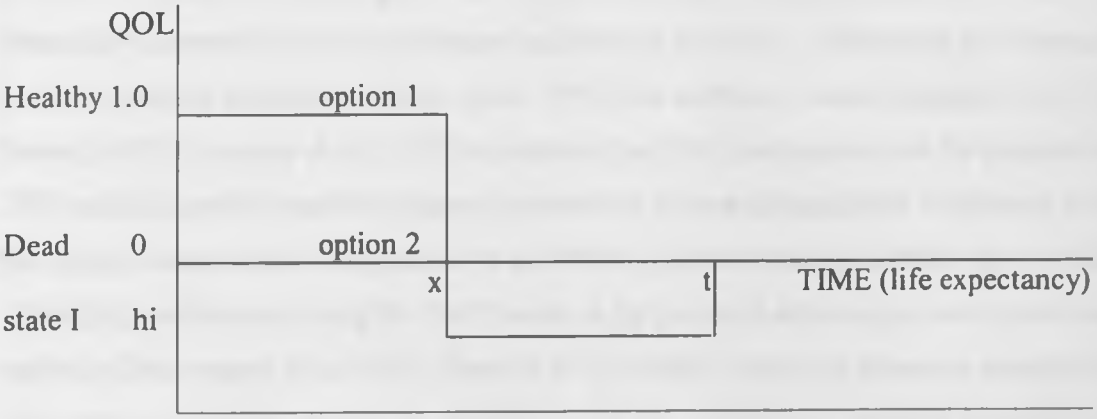


Figure 3f: TTO for a chronic health state considered worse than death



3.2.1.3 Rating Scales (RS)

The term rating scale covers both category rating or category scales and visual analog scale. Category rating is a simple partition method in which subjects assign each stimulus to one of a set of numbered categories (Kaplan and Ernest, 1983), say 1 to 10. Visual analog scale (VAS) consists of a line with clearly defined end points, usually 0 to 100, such that the stimulus can be rated on any point along that line. In RS, values are obtained in a risk-less and choice-less context. To obtain values, the most preferred health-state ("perfect health") is placed at one end and the least preferred ("death") at the other end to serve as anchors. The remaining health states are placed on the line (or category) between these two such that the intervals between the placements correspond to differences in preferences as perceived by the subject (Torrance, 1986; Froberg and Kane, 1989b; Torrance, 1987; Kaplan et al. 1993). Values can be obtained for chronic health states preferred to death, states worse than death and temporary health states (Torrance, 1986).

Because RS methods are choice-less, they are thought to be unable to measure strength of preference on a cardinal scale (Brazier et al. 1999b). Also the absence of opportunity cost implies that they have no basis in economic or decision theory (Bleichrodt et al. 1997) and thus have been regarded by economists as theoretically inferior to TTO and SG (Brazier et al. 1999b). Despite this, rating scales, founded on the theory of psychophysics (Kaplan and Ernest, 1983) specifically in psychometrics (Froberg and Kane, 1989b) are also claimed to produce cardinal measures (Kaplan et al. 1993). In addition, recent research (Dolan and Sutton, 1997; Torrance et al. 1996) has shown that VAS preferences can be mapped onto TTO and SG preferences using power conversion curves although the robustness of this link has not been established (Brazier et al. 1999b). However, this link offers the possibility of eliciting preferences using the VAS based on its practical advantages over choice based methods (Drummond et al. 1997; Brazier et al. 1999b), although whenever possible, the choice based methods should be preferred over the choice-less methods (Brazier et al. 1999b).

3.2.1.4 Person Trade Off (PTO)

The PTO was originally developed by Patrick et al. (1973) who called it the equivalence of numbers procedure (Richardson, 1994) but Nord (1992, 1994, 1995) called it PTO. In PTO, subjects are asked to decide how many people in health state B are equivalent to a specified number of people in health state A. The subject is asked, " If there are *x* people in adverse health situation A and *y* people in adverse health situation B, and if you can only help (cure) one group, which group would you chose?" One of the numbers *x* or *y* is varied until the subject finds the two groups equivalent in terms of needing or deserving help. The undesirability (disutility) of situation B is *x/y* times as great as that of A. For example, the task may be presented to the subjects as follows:

"*The first group contains 100 people in a state of maximum health (standard). Persons in the second group are in the state of health lower than the standard [specified] ..... How many people in this state of health do you consider equivalent to the 100 people of the same age in the standard group? You may use any number equal to or greater than 100*" (Froberg and Kane, 1989b. p.464).

Merits of the PTO are that it relates to social choice with outcomes relating to welfare of others, that it asks the right question (Pinto, 1997) and, that it is choice based. However, the method is argued to have no basis in economic and decision theory, because the opportunity cost is not borne by the individual. As such, consumer theory is not applicable (Brazier et al. 1999b). In addition to lacking a theoretical basis in economics, the PTO has practical difficulties in its use (Nord, 1995; Drummond et al. 1997; Murray and Lopez, 1997b; Sadana; 2002) and other than the GBD weights from Murray and Colleagues, it has not been widely used to value health states or with general populations (Brazier et al. 1999b).

3.2.1.5 Magnitude Estimation (ME)

In ME, subjects are asked to provide a number indicating how much better or worse each of the other states is compared with the standard. The task may be framed as follows:

"Let's give the first case the number 10. Now assign numbers to the other cases using the number 10 as your guide. For example, if a case seems 10 times as desirable as the first case, you would use a number 10 times as large or 100. If it seems one-fifth as desirable you would use the number 2 and so forth. Use fractions whole numbers or decimals, but make each assignment in relation to the desirability of the first case, as you see it." Froberg and Kane, (1989) P. 463. Magnitude estimation is a psychometric method that provides a direct estimate of the subjective ratio. The method is not based on any specific theory of measurement and gains credibility only through its face validity (Kaplan et al. 1993). ME is not choice based and hence like the RS, it lacks theoretical support in consumer theory or economics literature. Brazier et al. (1999b) noted that the ME is largely unused and is theoretically undeveloped, hence considered theoretically inferior by economists. However, it was the basis behind the Rosser and Kind matrix of values used in the first UK QALYs (Rosser and Kind, 1978).

3.2.2 Use and performance of valuation techniques in high income countries

This section briefly presents the use and performance[20] of valuation techniques described above based on reports from various studies. Table 3.2[21] shows that valuation techniques have been applied across countries and among different populations and disease groups. The most commonly applied techniques are SG, TTO and RS. Use of the PTO has been mainly in methodological environment (Brazier et al. 1999), but recently the GBD study has used it to elicit values for disease states (Murray, 1996; Fox-Rushby, 2002). The applications have concentrated mainly in North America and UK, where these methods were developed, with less application in other high-income countries. In terms of conditions, there is more application of the methods amongst chronic conditions and non-communicable conditions such as cancer and low birth weight. The tools have also been applied amongst different populations including members of the general population and university students and staff.

[20] The criteria for judging performance of both valuation and measurement instruments is presented in appendix 3.1.

[21] Studies included in tables 3.2-3.4 do not include all studies applying these techniques and is only meant to be indicative, as the review was not systematic.

Table 3.2: Application of Valuation Techniques in high-income countries[1]

	SG	TTO	RS	PTO	ME
Countries					
North America[a]	30, 17, 31, 32, 11, 14, 9	9, 2, 25	35, 19, 9, 31, 32, 11, 14		
United Kingdom	34, 3, 4,	8, 5, 1, 3, 4, 26	36, 29, 37, 34	36	42
Netherlands	33, 10,	10, 27,	38, 20, 33		
Switzerland		28			
New Zealand		7			
Spain	21		21,39	21	
Australia				23	
Norway				41	
Conditions / populations					
Cancer	32, 10	8, 10, 28, 5, 6, 7, 29	29, 32		
Hospitalized aged patients		25			
General population		1, 3, 4, 26,	34, 39		
Extremely Low Birth Weight	17				
Depressive illness	31		31		
Coronary Artery Bypass	14				
Students			33, 21, 19, 38		
physicians			14		
Nursing home residents			35		
University staff			36		

[1]Numbers in the table refer to the references listes below

1-Buckingham et al, 1996. 2-Mohide et al, 1988. 3-Dolan et al, 1996a. 4-Dolan et al, 1996b. 5-Ashby et al, 1994. 6-Perez et al, 1997. 7-Unic et al, 1998. 8-Buxton and Ashby, 1988. 9-Torrance et al, 1972. 10-Stigglebout et al, 1994. 11-Torrance, 1987. 12-Torrance, 1986. 13-Torrance, 1976. 14-Read et al, 1984. 15-Froberg and Kane, 1989b. 17-Saigal et al, 1999. 18-Kaplan and Ernest, 1983. 19-Patrick et al, 1973. 20-Essink Bot et al, 1990. 21-Pinto, 1997 22-Nord et al, 1999. 23-Richardson and Nord, 1997. 25-Tsevat et al, 1998. 26-Dolan and Gudex, 1995. 27-Stalmeir et al, 1996. 28-Hurny et al, 1998. 29-Johnston et al, 1998. 30-Llewellyn-Thomas et al, 1982. 31-Lenert et al, 1999. 32-Boyd et al, 1990. 33-Krabbe et al, 1997. 34-Dolan an Sutton, 1997. 35-Patrick et al, 1994. 36-Dolan and Green, 1998. 37-Robbinson et al, 1997. 38-Bleichrodt et al, 1997. 39-Badia et al, 1999. 41-Nord, 1992, 1994 and 1995. 42-Rosser and Kind, 1978.

a-North America includes studies done in both USA and Canada

Table 3.3 presents information on how the SG, TTO, RS, PTO and the ME have performed in terms of reliability, validity and practicality. Test retest reliability is the most common assessed form of reliability. While all methods can be considered to have satisfactory test-retest reliability, for the PTO it is reported to be low (Nord et al. 1999).

Table 3.3: Performance of Valuation Techniques in high-income countries by criteria

Criteria	SG	TTO	RS	PTO	ME
Reliability					
Test re-test	0.77[15 (1wk)], 0.49[15 (1yr)], 0.56-0.83[3] 0.85[14]	0.787[2], 0.81[3], 0.73[4] 0.5-0.67[5], 0.76[7], 0.96[7]	0.95[14], 0.77[15 (1wk)], 0.49[15 (1 yr)]	Low (testing on-going)[22]	0.74-0.83[15]
Split test	0.8-0.9[13]				
Intra-rater	0.7-0.94[15]		0.7-0.94[15]		
Inter-rater	0.75-0.77[15]		0.75-0.77[15]	0.6[15]	
Internal reliability	0.77-0.92[11]	0.77-0.88[11]	0.86-0.94[11]		
Precision[a]	0.13[11]	0.13[11]	0.09-0.15[11]		
Validity					
Convergent					
Construct		0.12-0.22[1]			
Concurrent	0.65[13] (TTO&SG), 0.36[13 (SG&RS)], 0.63[14(SG&RS)], 0.65[14 (TTO&RS)]	0.21-0.29[6]			
Predictive	<0.4[21]			<0.4-0.621[21]	
Criterion	Valid by definition[11,12]		Poor[11]		
Practicality					
Easy to administer*		Yes[2,5,8,9] (>SG)	Yes (>SG, TTO)[18,19]		
Response rate (%)	84-87[17]	54[1], >90[3]	57[20]	Low[23]	
Completion rates (%)		>90[2], 78[7]			
Completion time (min)		20-35[2], 50-74[7], 60-150[10], 20-25[8]			

1-Buckingham et al. 1996. 2-Mohide et al. 1988. 3-Dolan et al. 1996a. 4-Dolan et al. 1996b. 5-Ashby et al. 1994. 6-Perez et al. 1997. 7-Unic et al. 1998. 8-Buxton and Ashby, 1988. 9-Torrance et al. 1972. 10-Stigglebout et al. 1994. 11-Torrance, 1987. 12-Torrance, 1986. 13-Torrance, 1976. 14-Read et al. 1984. 15-Froberg and Kane, 1989b. 17-Saigal et al. 1999. 18-Kaplan and Ernest, 1983. 19-Patrick et al 1973. 20-Essink Bot et al. 1990. 21-Pinto, 1997 22-Nord et al. 1999. 23-Richardson and Nord, 1997.

* >=easier than. <=harder than

[a]- precision was expressed as Standard Deviation

In general, test retest reliability is low for longer periods changing from 0.77 for 1 week to 0.49 in one year (Froberg and Kane, 1989b). Brazier et al. (1999b) build on previous reviews and report similar additional findings as shown in table 3.4. A general lack of evidence surrounding the reliability of methods exists (Brazier et al. 1999) especially for the PTO and ME.

Table 3.4: Test retest reliability of the SG, TTO and VAS techniques

Test retest reliability	SG	TTO	VAS
1 week or less	0.80[a] 0.77-0.79[j]	0.87[a]	0.77[a] 0.70-0.95[j]
4 weeks	0.82[b]	0.81[c] 0.63[d]	0.62[b] 0.89[d]
3-6 weeks		0.50-0.75[l]	
6 weeks		0.63-0.80[c] 0.85[e]	
10 weeks		0.73[f]	0.78[g]
6-16 weeks	0.63 (props)[m] 0.74 (no props)[m]	0.83 (props)[m] 0.55 (no props)[m]	
1 year	0.53[h]	0.62[h]	0.49[h]
Other (time unspecified)	0.82[i] 0.80[k]	0.74[i] 0.67-0.92[k]	

[a]- O'Connor et al. (1985), [b] - O'Brien and Viramontes (1994), [c] - Churchill et al. (1987), [d] - Gabriel et al. (1994), [e] - Molzahn (1996), [f] - Dolan et al. (1996a), [g] - Gudex et al. (1996), [h] - Torrance et al. (1976), [i] -Read et al. (1993), [j] - Baker et al (1994), [k] - Gage et al. (1996), [l] - Ashby et al. (1996), [m] - Dolan et al. (1996b).
Correlations undertaken where specified: Interclass correlation coefficient – b,f,g,j,k; Pearson correlation coefficient – d,m; others unspecified.
Reliability for the PTO and ME is yet to be demonstrated but is comparatively low
Source: Brazier et al. 1999b. p.31.

There is paucity of literature reporting on the validity of valuation techniques (Brazier et al. 1999b). Table 3.3 shows that amongst the few assessing validity, most studies have assessed concurrent validity by applying SG, TTO and RS together in various combinations and have found correlations ranging between poor (0.21-0.29) (Perez et al. (1997) to satisfactory (0.65) (Torrance, 1976 and Read et al. 1984). A notable finding in the literature is that the RS do not correlate well with choice based methods such TTO and SG (Torrance, 1976; Read et al. 1984; Brazier et al. 1999b), while the choice-base methods correlate reasonably well.

Table 3.3 shows that in terms of practicality, RS is easier to administer than TTO and SG. The RS methods have greater level of completion when compared with the TTO and SG (Brazier et al. 1999b) with response rates varying between 57% for postal questionnaire to over 90% for administered questionnaire (Buckingham et al. 1996; Mohide et al. 1988; Saigal et al. 1999; Essink Bot et al. 1990). Although Froberg and Kane (1989b) considered

the SG as complex for population studies and expensive to undertake, it has been used successfully in population studies (Dolan et al. 1996a). The TTO and SG perform more or less similarly in terms of practicality (Brazier et al. 1999b). Although reliability, validity and practicality are found to be satisfactory for SG, TTO and RS in settings where the instruments were developed, mainly North America and Europe, they vary widely in their performance. This warrants fresh assessment of performance of the techniques whenever used in new settings such as developing countries where little is known about their performance.

PTO and ME have been studied less. The reliability, validity, feasibility and acceptability of the PTO are relatively unknown (Nord, 1995; Brazier et al. 1999a; Brazier et al. 1999b; Green et al. 2000). Its' earlier version (equivalent technique) was reported to have confused and offended respondents (Patrick et al. 1973 *cited in* Brazier et al. 1999b) and Froberg and Kane (1989b) note that it has been considered offensive to respondents. It is also quite demanding (Nord (1995) as evidenced in the GBD study PTO protocol that have lasted 10 hours (Murray and Lopez, 1997b; Sadana, 2002). Respondents also require a lot of motivation to stay involved and the response rates are very low (Richardson and Nord, 1997). In their review of valuation techniques, Brazier et al. (1999b) found no evidence to demonstrate acceptability and validity of ME although it was reported to have reliability in Froberg and Kane (1989b). Due the paucity of evidence on their performance in their original settings (in terms of psychometric properties and practical considerations in application), it would be pre-mature to consider transferring the PTO and ME to another setting. They will not be considered for application in this study.

3.2.3 Understanding Variation in Values

Consumer theory points to a number of factors that may affect values generated in a preference elicitation task. These include human characteristics, attitudes and behavior (rationality, consistency and transitivity), wealth or income, socio-cultural, value and belief systems, characteristics of the goods among which a choice is made, as well as amount information about the good and consequences of the choice among others. Preferences may

be affected by whether certainty[22] or uncertainty[23] prevails during the choice process. Therefore an individuals preferences for health states could vary depending on their attitudes to risk and these attitudes have been shown to vary depending on whether prospects are framed as gains or losses (Kahnemann and Tversky, 1979). Human characteristics such as life stage, age, sex, level of income and education, tastes and preferences and family circumstances could also cause variation in values for health states.

According to consumer theory, we cannot predict how age and sex would affect values. However, if all individuals, irrespective of their personal characteristics, are rational, consistent and transitive, then we would expect their values to be lower the worse off health states are and the longer the duration of health states. Glahe and Lee (1981: p. 72) point out that the curvature of the indifference curve implies that, the smaller the amount of good 1 held and the larger the amount of good 2 held, the more valuable are the marginal changes to the consumer in good 1 relative to marginal changes in good 2. Hence, the consumer will be willing to give up larger amounts of the good in abundance to obtain a unit of the other scarcer good. In relation to health states this can be interpreted as inferring that the worse off a health state is, the more years one would be willing to give up to be healthy, hence the lower the value. Wealth or income and education are positively related. From the consumer theory, positive or negative income effect (either due to change in income or change in prices) depends on whether a good is considered as a normal or inferior good. Assuming that health is a normal good, then the theory predicts that the higher the income the higher the values assigned to health states. However, other factors may affect this prediction, such as value and belief systems, attitudes, psychological factors affecting people's behavior as well as amount of information about the good and the consequences of a given choice.

As mentioned earlier values have been found to vary across groups of valuers i.e. patient, health professional or general population, and within methods (Drummond et al. 1997;

[22] Choice under certainty implies that consumers know for sure the consequences of their choices and there are no risks involved. Both ordinal (indifference curves theory and revealed preference theory) and cardinal theory have been used to study consumer choice under certainty (see chapter 2).

[23] Choice under uncertainty implies that there are probabilities attached to the outcomes of the choices made (risk) or there are no such probabilities (as in states of nature whose probabilities are not known). von Neumann-Morgenstern expected utility theory is applied in the study of choice under uncertainty (See chapter 2). This theory introduced a new form of cardinality in utility assessment (O'Brien, 1990).

Brooks, 1995; Gold et al. 1996). It has been argued that it may not matter whose preferences are used if it can be demonstrated that no major differences exist among groups of raters (Froberg and Kane, 1989c). Consumer theory (which assumes certainty) emphasizes consumer sovereignty and would therefore favour values from patients. However, uncertainty prevails in terms of risk of being potential patients and success of treatments. Since members of the general population are the potential patients, their values are relevant. Buxton and Ashby (1988) and Torrance (1986) argue that if the choices involved are for the care of an individual, then that individual's value are relevant (consumer sovereignty) but if the decision to be made concerns a group, then group values are relevant. Values from the general public have been recommended on the basis that it is the society's resources that are being allocated (Torrance, 1986; Gold et al. 1996).

Table 3.5 shows a range of factors that may cause variation in values from various studies. The range of factors include age, sex, educational level, risk attitudes, experience with illness and type of valuer, duration of state, time horizons used in valuation, framing of the questions and severity of the health state. There is lack of agreement on how age, sex, and educational level affect values, with some studies finding positive and others negative effects. With respect to experience with illness, those with dysfunction give higher scores. However, definition of "experience with illness" should be explicit as it appears that being in ill health, knowledge of health state, past experience with the illness, appear to have differing and conflicting effects. The valuation of a given health state is a decreasing function of both its duration and severity. Other factors such as level of income or wealth, socio-economic status, perceptions of the disease (its importance in relation to other diseases) have received no attention although from consumer theory they may affect values. These factors have not been studied in developing countries, hence the justification for their study to facilitate the choice and application of the existing health-state measurement and valuation techniques.

Table 3.5: Factors Affecting Values for health states

Factor	Effects
Age	• Elderly consider health improvements less important [1] and demonstrate a strong will to live rather than trade off quantity for quality[10] • Older respondents produce significantly higher numbers of inconsistencies[15,16,19,17,21] • Inconsistency rates are positively related to age[10],[15,19] • The younger give higher valuations[3] • Age, gender, employment status show no systematic influence on valuations[2,5,6,9,11]
Sex	• Values affected by age and sex, both statistically significant[3] • No relation between age, sex, educational level or race[5,6,10,11] • Men give higher valuations than women[3,16,17]
Educational level	• Respondents with lower levels of education produce significantly higher numbers of inconsistencies[15,19,23] • Inconsistency rates are negatively related to educational attainment[19]
Risk attitudes (gambling effect)	• Respondents risk averse[6] • Reluctance to gamble with own health[14,21]
Experience with illness and type of valuer	• Values from patients differ from non-patient groups[2,3,4,9,15,16,17,13,19] • Own health state influences value assigned to a state in that the ill assign higher values than the non-ill[18] • Direct knowledge of what life is like with a colostomy affected values (gives higher values) [13,7] • Those in poorer health generally give higher valuations[16] • Past experience with illness appears to have negligible effect on valuations[16] • valuations from patients with a particular illness are often higher than those from patients without the illness[5] • Those in a dysfunctional state give significantly higher scores[2] • Values of chronically ill individuals differ significantly from those obtained from healthier individuals, particularly in the case of more severe health states[18]. Health individual assign negative values to some states while the ill rate all positively[18] • Patient s with chronic illness systematically assign higher values to all health states[13,15] • Doctors in general place a lower valuation on physical components of ill health[4,5]
Duration of health state	• Utility for a given state depends on quantity, time and gambling effect[13, 14] • The measured utility of states of ill health declines with increasing time spent in those states[12] • Poor states of health become more intolerable the longer they last. The valuation given to a health state is a decreasing function of both its severity and its duration[16] • The worse a health state is the more willing people are to trade off years reflecting some "threshol of tolerability" before years of life can be given up[20] • The differences in values between RS and TTO can be explained by the fact that in RS, there is a tendency for people to ignore the duration of the health state, and which is clearly specified in TTO[20]
Time horizons in health states	• Different time horizons used in valuation exercise result in different values[7,8]
Framing of the questions	• People have difficult accepting death as a health state[18]
Severity of health state	• The more severe a health state is the less the utility value assigned to it[12,18]

1-Buckingham et al. 1996. 2-Dolan et al. 1996a. 3-Dolan et al 1996b. 4-Ashby et al. 1994. 5-Buxton and Ashby, 1988. 6-Stigglebout et al. 1994. 7-Torrance, 1987. 8-Read et al. 1984. 9-Froberg and Kane, 1989b. 10-Tsevat et al. 1998. 11-Johnston et al. 1998. 12-Llewellyn-Thomas et al. 1982. 13-Boyd et al. 1990. 14-Bleichrodt et al. 1997. 15-Badia et al. 1999. 16-Dolan, 1996. 17-Gudex et al 1997. 18- Badia et al. 1998. 19-Dolan and Kind, 1996. 20-Robinson et al. 1997. 21-Nord, 1992

3.2.4 Cross cultural issues in Valuation

Valuation of HRQL in developing countries is a relatively recent phenomenon spanning 1991-2002, with only a handful of studies testing the application of valuation techniques amongst low-income populations. Table 3.6 shows the application of valuation tools in low-income countries. The countries represented include the Philippines, Kenya, Zimbabwe, Burkina Faso, Cambodia and India. The VAS has been the most commonly used tool. Although Sadana (2002) found it impossible to use the PTO amongst the general population in Cambodia, Murray and Lopez (1996) have used it extensively in the computation of DALYs for priority setting in developing countries (Fox-Rushby, 2002). In all the applications, the tools have been used to value disease states with the exception of Jelsma (2002) who used the EQ-5D health states. With the exception of Tan Torres (1991) all other studies have used the general population, in addition to either patients or health professionals to provide preferences. Nearly all the studies were engaged in testing the methodologies and whether the instruments could be used with these populations. While this is a reflection of the novelty of the use of valuation of health states in developing countries it also points to the need to account for cross-cultural issues in transferring these instruments to new settings and subsequent testing of validity, reliability and practicality.

Out of the six studies identified that have used valuation tools in developing countries, two (Sadana, 2002 and Mahapatra et al. 2002) have not explicitly reported on the validity and reliability of the tools although they allude that the tools perform reasonably well. Table 3.7 shows satisfactory test-retest reliability for the VAS in the Philippines, Kenya and India and for the TTO and SG in the Philippines. The VAS also has good convergent and construct validity, but poor concurrent validity. Although the studies report findings from first use of the instruments in these settings, it is clear from table 3.7 that the reporting and or assessment of validity, reliability and practicality has been poor.

Table 3.6: Use of valuation tools in low-income countries

Author/year/country	Valuation tool used	Disease states valued	Whose values
Tan Torres, 1991/ Philippines	VAS, TTO, SG	Leprosy	patients
Kirigia, 1994 / Kenya	VAS	*S. Mansoni*	General population, teachers, health professionals
Jelsma, 2002 *cited in* Brooks et al. 2003 /Zimbabwe n=2182	TTO	EQ-5D states	Community members
Baltussen et al. 2002 / Burkina Faso / N=56	VAS	9 hypothetical GDB states	Lay individuals and health professionals
Sadana, 2003 / Cambodia / N=40	VAS	11 of the 22 GDB indicator conditions and 15 reproductive health and illness indicator conditions	Non-health professional women from community (n=20) and seeking care (n=20)
Mahapatra et al. 2002 / India / n=1010	VAS	22 GBD states	Community members
Murray and Lopez, 1996	PTO	GBD conditions	Experts

Table 3.7: Performance of Valuation Techniques in low-income countries by criteria

Criteria	SG	TTO	RS	PTO	ME
Reliability					
Test re-test	0.49-0.61[1]	0.54-0.82[1]	0.54-0.82[1], 0.90[2]		
Internal reliability			0.6-0.8[5]		
Validity					
Convergent			Good[1]		
Construct			0.86-0.94[2]		
Concurrent			Poor[1]		
Practicality					
Response rate (%)			>90[1,2]		
Completion time (min)	8-12[1]	8-12[1]	8-12[1], 12[3]		
Comprehension rates[a] (%)	41[1]	51[1]	31[1]		
Mean comprehension rates[b]	0.73[1]	0.7[1]	0.76[1]		

1-TanTorres, 1991. 2-Kirigia, 1994 3-Baltussen et al. 2002. 4-Sadana, 2002. 5-Mahahpatra et al. 2002. 6-Jelsma, 2002.

* >=easier than. <=harder than

a- Subject comprehension was graded by research assistants using a visual analog scale with 1=complete comprehension and 0=no comprehension at all (Tan-Torres, 1991). [b]omprehension rate is the proportion comprehending the method.

NB// Sadana, (2002) has not provided any data on how the VAS performed

Tan Torres (1991) used patients and caregivers to construct scenarios[24], which were subsequently valued in an attempt to maximize content validity. Kirigia (1994, 1997,1998a, 1998b) used a modified version of the RS, the rice-sack visual analog scale to elicit preferences for various schistosomiasis health state scenarios in Kenya. Unlike Tan-Torres, Kirigia used one schistosomiasis epidemiologist to construct health state scenarios and failed to take into consideration the views of the community that experience the illness. He constructed seven severity stages in *S. Mansoni* and accompanying clinical symptoms and their impact on victims six functional dimensions namely; mobility, self-care, livelihood, energy, pain and social participation. Kirigia (1994) used VAS on the basis of its ease of application and dismissed all the other techniques on the argument that they cannot be applied with illiterate populations. However, Kirigia (1994) did not establish validity and reliability of the VAS, although the overall response rate was 99% compared with 100% in Tan-Torres (1991). Kirigia (1998) suggested that; a population survey is necessary to establish relevance or irrelevance of functioning and ill-health among the community studied; use of mock exercise to enhance and familiarize respondents with the instruments; that to obviate re-inventing the wheel test the relevance of key HRQL instruments currently being used in developed countries while acknowledging the differences in language, culture and setting.

Recently, Baltussen et al. (2002) used a modified version of the VAS in rural Burkina Faso to value nine hypothetical disease states (BOD) relevant health states. Other recent applications include the use of the VAS to value disease states in Cambodia amongst reproductive aged women (Sadana, 2002) and in India in a community survey to value a variety of disease states (Mahapatra et al. 2002) including those of the GBD study by Murray and Lopez (1996). Amongst these researchers, except for Baltussen et al. (2002) none considered whether the instruments would be equivalent in the new settings and therefore cross-cultural issues have been ignored. Although, Kirigia modified the VAS to the rice-sack VAS, it is not exactly clear what led to this modification although he cited the

[24] To construct scenarios, Tan-Torres (1991) interviewed 23 patients, caregivers and health professionals. Questions centered on impact of disease on 6 parameters namely, physical mobility, activities of daily living, degree of social interaction, emotional status, pain and productivity. Patients were asked to describe in their own words their status regarding these parameters both before and after treatment.

low levels of education. Jelsma (2002) *cited in* Brooks et al. (2003) provides details of how the EQ-5D instrument was adapted and translated into Shona, but not the TTO. In general therefore, issues arising when valuation instruments are used in settings in which they were not developed have not been adequately addressed and this constitutes a knowledge gap.

Fox-Rushby et al. (2001) and Mugo and Fox-Rushby (2003) have considered the relevance of health state valuation techniques in Kenya, specifically examining issues of conceptual equivalence, in recognition of the cultural differences that might exist between the original and target settings. Both found evidence of cultural differences especially with regard to issues of how death, conceptualization of perfect health and willingness to trade are considered in the instruments. They emphasise the need for local understanding of concepts embodied in instruments to facilitate modifications and enhance the potential for comparisons of values when instruments are used in different settings. Assessment of equivalence leads to an understanding of how different terms and concepts embodied in the instruments are conceptualized and how important they are considered in the local worldviews which could affect values in different settings. This phenomenon has been shown in Brooks et al. (2003), Fox-Rushby and Selai (2003), Herdman et al. (2003) and Rabin et al. (2003). For example in Zimbabwe, being in severe depression and being confined to bed were regarded as less serious in the Zimbabwean sample while self-care level 3 was considered more serious in Zimbabwe than the UK (Brooks et al. 2003). Health states including these levels in these two settings are likely to be valued quite differently. While this example related to domains in EQ-5D, the principle would possibly apply in cross-cultural use of valuation instruments.

In establishing the health states to be valued, neither Kirigia nor Tan-Torres used any of the existing HRQL health status classification instruments. Therefore, from their instruments we cannot tell the spectrum of health states from these illnesses, as they are restricted to the states they valued. Knowing the breadth of health states resulting from a disease and desirability of moving from one state to another is integral not only to measurement but valuation of health.

3.3. CONCLUSION

This chapter has presented a range of methodological and conceptual issues with regard to measurement and valuation of health/disease states. Special attention was paid to use of HRQL tools in developing countries for health care decision making. It was noted that measurement of health related quality of life has largely been undertaken in the developed countries, mainly North America and Europe where most of the instruments are developed. The focus has also been on chronic conditions like cancer as well as diseases associated with an ageing population. Although problems like cancer and concerns for elderly people may be present in developing countries, preventable and treatable infectious and parasitic diseases affect the majority of the population and are main causes of morbidity and mortality. Conditions such as *S. Mansoni* have therefore not been studied and warrant researching.

Interest in applying HRQL measures in developing countries is new (Bowden and Fox-Rushby, 2003) with no application of measurement instruments in the Kenyan setting. The task of developing a generic measure requires considerable time, resources and multi-disciplinary expertise that go beyond the scope and focus of this thesis. It was noted that except for Jelsma (2002) cited in Brooks et al. (2003) who used the EQ-5D in Zimbabwe, other studies have used the GBD study indicator conditions or had the authors define the disease states with the help of patients or experts. They have therefore provided no way of measuring or ascertaining what other disease states a different population may experience, other than the rigid ones they construct. This therefore calls for development of a disease specific measure that can be used in repeated studies with different populations.

Due to the enormity of developing and or adapting a generic measure into new settings and the strict requirements by the original instrument developers, the empirical study in this thesis has focused on developing a disease specific instrument to assess impact on HRQL. It is noted that health is a broad concept ranging from the narrow biomedical model of health focusing on disease through consequences of disease to broad conceptions that include external, environmental, cultural and social influences on the person. Owing to the

newness of application of these concepts and knowledge in Kenya, this work focuses on a specific disease and assesses the consequences of disease on people's health related quality of life.

Although measurement of health and disease states is useful in its own right, economists require that measured disease/health states be valued to allow for efficiency evaluation. The need to separate measurement and valuation of health/disease states was underscored, and is reflected in this thesis. Five non-monetary valuation techniques were presented together with their theoretical basis and performance in terms of validity reliability and acceptability. On theoretical grounds, the SG is the preferred technique among economists. However, this theoretical base has been challenged and therefore, SG should be evaluated along other criteria and on the same basis as other methods (Geisler et al. 1999). Based on the criteria of validity, reliability and feasibility and acceptance, the SG performs similarly to the TTO (Torrance, 1987: Brazier et al. 1999b; See table 3.1). Their performance is regarded as satisfactory (Bowling, 1997) but it varies widely. This implies that in deciding to adopt TTO as a substitute for SG as suggested in the literature (Torrance, 1972, 1987), their performance, in terms of validity, reliability and practicality, and equivalence in a given setting should be established first. In addition, the factors that influence valuations in such a setting should also be established. It was established in the review that the SG and TTO have been inadequately studied in the developing countries and their application in a tropical parasitic disease such as schistosomiasis is not known.

Froberg and Kane, (1989d) report that RS has been the most well researched method followed by TTO and SG. Although the method is easy to administer, is cheap and efficient for application in large surveys, both postal and administered, it suffers from lack of a theoretical basis in decision theory for which the values are required. However, while some argue that the method produces preferences on an interval scale (Brooks, 1995), others suggest that due to response spreading this property may not present (Brazier et al. 1999b). It performs equally well on reliability as SG and TTO, although its validity is inconclusive (See table 3.2). As this method appears most feasible and easiest to understand, it may be easily applied with non-literate populations. However, its performance alongside other

instruments needs to be established and judged in such situations so as to make a justified choice for application in a developing country setting.

A range of methodological and conceptual issues in valuation of health was presented. Due to their diversity, their study would require different study designs and therefore they cannot all be addressed in one study. However, in recognition of the novelty of application of valuation tools in developing countries, it is clear that methodological issues of validity, reliability and practicality deserve priority in any new setting.

Although it has been argued that the developing countries population are largely illiterate (Kirigia, 1994) and therefore cannot handle the TTO and SG tasks, this claim is unsubstantiated in that the technique has not been tested in those circumstances and found infeasible and or unacceptable. Besides, literacy defined as not being able to read and write, does not necessarily imply that the cognitive abilities of these populations are compromised and that they therefore cannot reason and comprehend the task. This is especially in an administered interview and with use of visual aids such as the probability wheel and chance board accompanied by warm up or mock exercise to familiarize subjects with the valuation tasks. Thus far, the justification provided for not using TTO and SG as well as existing HRQL instruments in developing country is not adequate[25] and needs further testing, particularly to understand reasons for or against using such techniques obtained from populations where it is to be employed.

A variety of factors affect preferences and different methods of preference elicitation produce different values. Since the question of whose values to use in decision making is still unresolved and not known in developing countries, it warrants research. This will provide informed decisions on whose values to consider at various levels of decision making. Variation in values also occur within an individual because of differences in tastes and preferences, risk attitudes, personal socio-demographic and economic characteristics, their perceptions of the goods among others. Studying these factors gives insights into how

[25] These instruments have not actually been put to test in those settings but have been dismissed based on findings of studies in other settings. More evidence is therefore necessary.

well theory can predict choices and whether and how different groups of individuals violate or confirm the axioms upon which consumer choice is based. They therefore provide important information into validity of the valuation techniques. The empirical study of this thesis has therefore assessed how various characteristics of the rater as well as of the disease states influence values.

Methodological issues of validity, reliability, practicality and equivalence need to be investigated for both measurement and valuation instruments both when newly developed or if being adapted for application in new settings. As no health outcome measurement instrument specially developed and or adapted for use in a Kenyan setting or for the specific disease being investigated in this study was identified, a new tool for assessing consequences of *S. Mansoni* was developed. Similarly amongst valuation techniques, very little application of the common and psychometrically sound techniques has been undertaken in Kenya. Further testing of applicability of these methods in a new setting requires adapting them and testing for their validity, reliability and practicality to determine their suitability for use in Kenya.

CHAPTER 4

A SYSTEMATIC REVIEW OF DISEASE SPECIFIC VALUES/UTILITIES AND VALUATION OF DISEASE-SPECIFIC STATES

4.1 INTRODUCTION

This chapter presents a critical literature review of the current state of affairs in eliciting disease specific utilities. It focuses on methodological and empirical issues relating to disease specific utilities and how authors view their use. Special attention will be paid to use of disease specific utilities in economic evaluation in low-income countries.

In terms of methodology, the review will determine how issues of validity, reliability, practicality and equivalence have or have not been addressed in relation to disease specific measurement and valuation. The review will also determine the extent to which measurement and valuation instruments are applied in settings other than those in which they were developed and how issues of cross-cultural adaptation have been addressed to date. By highlighting all the possible methodological and empirical gaps in knowledge, this chapter will contribute to justifying the framework and issues for study in the empirical chapters of the thesis. A discussion of the main issues and conclusions on the way forward will be presented.

The specific objectives of the review are to address

1. Which populations were studied, in terms of subject, diseases or conditions and countries of study for both measurement and valuation;
2. The objectives of measuring and valuing disease specific utilities presented in the studies and how they relate to economic evaluation and health care decision-making;
3. The stated rationale for the valuation methods used and how this relates to the theoretical framework of economic evaluation;

4. How validity, reliability, practicality and equivalence are conceptualized and determined, with respect to measurement and valuation instruments;
5. What is measured and valued;
6. How are the disease states derived and justified; the methods and processes used to construct the utility measure;
7. Who does the valuation and why?
8. How issues of cross-cultural adaptation arise and addressed;

The chapter consists of four sections. Section two presents the methods used to search, select and review the literature. Section three presents results of the review. It begins with presentation of the background details, followed by findings on methodological and empirical issues in assessing disease specific utilities separately for valuation and measurement instruments. Section four discusses issues outlined above, highlights current gaps in knowledge and selects those to be addressed further in the thesis.

4.2 REVIEW METHODS

4.2.1 Scope of literature search

For the purpose of this review, disease specific utilities will be considered as utilities derived for any disease specific scenario using any of the existing non-monetary valuation techniques (Rating scale (RS) time trade off (TTO), Standard Gamble (SG) person trade off (PTO) and Magnitude estimation (ME)). Only primary research studies using disease specific instruments to generate disease specific scenarios and/or disease/health classifications, which are valued using any of the existing valuation techniques are identified.

4.2.2 Literature Search Strategy

Medline and PubMed database were searched from January 1966 to December 2001. To capture the types of papers needed, the search covered three broad areas: disease (condition) specific instruments (outcomes or measures); utilities, preferences and values; and valuation techniques and health state valuation. The search terms included

patient*, disease*, condition* each combined with measur*, instrument*, outcome*, using 'and' and 'or'. To include all diseases, PubMed MeSH Browser was used to obtain all MeSH categories for diseases categories. A total of 22 disease categories were identified (Appendix 4.1). These categories were used on Medline thesaurus and exploded using the broad term 'economics'. The terms utilit*, valu*, preference*, scal*, were searched and each combined with patient*, disease*, condition*. To capture health states valuation and valuation instruments, the following terms were used. Health stat*, technique*, TTO, VAS, ME, PTO, SG (as well as the full names and different forms of spelling and hyphenation eg. Time trade off and time-trade off). In addition, MeSH terms 'quality-adjusted-Life-years','quality-of-life','health-care-economics-and-organisations', 'health-status',health-status-indicators',and'outcome-assessment-health-care'were exploded using all subheadings to capture health measurement instruments. These terms were combined using the Boolean operators 'and' and 'or' and thesaurus function. Specific authors known to have done work in this area were searched using the index function of Medline[26]. One paper published in 2002 was identified later[27] and having met the review criteria was included for the review.

4.2.3 Inclusion and Exclusion Criteria

Inclusion

Papers meeting at least one the following criteria are included for review.

- Papers focused on primary research with an aspect of utility assessment or assessment of values or preferences for health (disease) states.
- Papers where the objectives of the study related to economic evaluation in the context of valuation and measurement of outcomes, development of disease/condition specific utilities as well as evaluation of methodological aspects of the instruments (psychometrics etc).
- Papers comparing use of disease specific utilities and or measures with generic or other measures.

[26] These included Ashby, J., Bombardier, C and Brazier, J.
[27] This paper was brought to my attention by a colleague in the Health Policy Unit, LSHTM.

Exclusion

- Papers based on modeling were excluded if utility assessment was not carried out in a population survey (i.e. if utility assessment was through modeling and estimations).
- Papers without abstracts.
- Papers on valuation but not focusing on disease states or using diseases specific instruments (e.g. those on valuation but using generic measures).

4.2.4 RESULTS OF THE SEARCH

The search yielded a total of 340 relevant papers. The first step of selection involved a title review to exclude papers unrelated to health outcome assessment. Papers bearing no relation to valuation (n=260) were dropped resulting in 80 papers whose abstracts were reviewed further based on the inclusion and exclusion criteria outlined above. This resulted in a further rejection of 39 papers, leaving 41, plus one identified for 2002, making 42 upon which this critical review is based. Of those rejected, 14 were not based on primary valuation research as utilities were assigned through expert judgement by the authors or professionals (n=2) and from previous studies (n=12). In six studies valuations were not for disease states while another six studies were rejected as they were based on modeling and it was not clear how the utilities had been obtained. Two papers were redundant in that their valuation was similar to studies already included in the review and the remaining 11 were not about valuation.

4.2.5 Systematic review procedure

All papers selected for review were subjected to the same set of review questions to ensure standardization and consistency. The review questions (see appendix 4.2) were categorized into those seeking background details and those concerning methodological and empirical issues and set out in about 17 tables. After reading each paper, the author went through each of the questions in the tables seeking to find out if the paper had addressed the issues and if so, how and what were the empirical findings (where relevant). Organizing the questions into tables allowed a systematic and simplified classification of information for ease of ana lysis. Having subjected all the papers to the

review questions, analysis consisted of frequency distributions and tallies. Therefore, reporting of results is in terms of absolute numbers and percentages, with further reporting of information as text.

4.3 REVIEW RESULTS

4.3.1 Background details

Forty-two studies reviewed spanned the period 1986 to 2002 (see appendix 4.3). The majority had been carried out in 1999 (n=10) and 2000 (n=10). Six of the studies had been undertaken in 1998 and the rest spread between 1986-1997 (n=13) and 2001-2002 (n=3). Most research has been carried out in Canada and USA, both accounting for 73.8% (31/42) of the studies. Both the UK and Netherlands each account for 4 studies followed by one in Kenya and one in Burkina Faso. One study was carried out in both North America and Europe.

To avoid confusion, the rest of this section reports results for studies that used measurement and valuation instruments alternatively. All studies reviewed used valuation instruments. Of these, 18 also incorporated both disease specific and generic measurement instruments. There is much diversity in the conditions studied. Rheumatic conditions, visual loss and breast cancer each had more than two studies focusing on measurement of the states. Each of the following conditions only had one study; diabetes mellitus, dental implants, angina, intermittent claudication, peripheral vascular disease, heart failure, chronic airflow limitation, osteoporosis, asthma, HIV infection and states of patients attending primary health care. Of the studies focusing solely on valuation, nine were in breast cancer and five in rheumatic conditions. Diabetes mellitus, and prostate cancer each had three studies while depression, visual loss, peripheral vascular disease, heart failure and psoriasis for two studies each. The rest of the conditions accounted for one study each and they include dental implants, angina, intermittent claudication, chronic air flow, osteoporosis, asthma, HIV infection, states of patients visiting primary health care, schistosomiasis, erectile dysfunction, rectal cancer and selected diseases from the Global Burden of Disease 22 indicator conditions. Of the 18 measurement studies 16 used patients as the respondents while 2 used both patients

and general public. Similarly, 25 out of 42 valuation studies used patients, 9 general public and 8 both patients and general public as study subjects.

Table 4.1 shows the sample sizes by measurement and valuation studies. The size of samples range from 13 to 528 subjects with one outlier (Sherbourne et al. 1999) having 17,707 subjects. Twelve of 18 measurement studies have sample sizes less than 100. Similarly, more than half (n=23) of the valuation studies had sample sizes less than 100.

Table 4.1: Study sample size

Range of sample size	Measurement studies	Valuation studies
1-50	4	11
51-100	7	12
101-150	1	8
151-200		2
201-300	2	2
301-400	3	3
401-500		2
Over 500	1	2
Total	**18**	**42**

4.3.2 Objectives and justification of studies

Table 4.2 shows the nature of objectives by measurement and valuation studies. Objectives not related to empirical or methodological work were classified as general objectives[28]. Twenty-four studies had methodological objectives while 16 had empirical and 15 general objectives. None of the general objectives in the 15 studies were related to measurement of health states. They all focused on valuation. The majority of studies had objectives related to obtaining values and utilities for use in cost utility analysis, while others aimed to use utilities to ascertain quality of life. More than a third of the 16 studies stating empirical objectives focused on analysis of utilities. A quarter of the studies assessed factors affecting preferences for health/disease states and relationships between disease states as well as generic states relating to specific conditions and quality of life as measured by utilities in each case. The majority of methodological objectives were focused on valuation. Out of 24 studies stating methodological

[28] Though some of these objectives are seemingly empirical, they were not concerned with directly investigating methodological and empirical issues in eliciting disease specific utilities

objectives 12 were focused on validity assessment, six on reliability, four on feasibility and four on comparisons of valuations between different groups of raters.

Table 4.2: Nature of objectives

Nature of objective	Measurement studies	Valuation studies
General objectives (n=15)[a]		
To elicit preferences for economic evaluation		9
To ascertain quality of life using utilities		4
Utility awareness creation in a discipline		1
To estimate QALYs for policy making		1
Empirical objectives (n=16)		
To measure utilities and perform utility analysis		6
To provide information on differences in preferences by various factors or characteristics	1	4
To determine relationship between disease states and quality of life as measured by utilities and quality of life instruments	4	3
To determine differences in preferences for states between methods	1	3
To determine differences in preferences for states between raters		2
To test QALY assumptions		1
Methodological Objectives (n=24)		
To assess validity of the instruments	1	12
To assess reliability		6
To explore feasibility of instrument		4
To compare valuations from different raters		4
To demonstrate development of a measure and to describe its design	3	3
To measure utilities		2
To evaluate responsiveness and sensitivity to change		2
Others		5

a Totals do not add up to sample size because categories of objectives were not mutually exclusive.

Table 4.3 presents the reasons provided as justification for undertaking various studies. The most commonly provided justification for undertaking research on disease specific utility elicitation was that utility measurement has not been applied to the condition before, or that very little research on utilities has been done in the disease area. Another justification was the increasing recognition of the importance of incorporating quality of life concerns into the assessment of outcomes of interventions. Other rationales included facilitating economic evaluation of treatments.

Table 4.3: Justification for undertaking studies

Justification	Measurement studies	Valuation studies
Utility measures have not been applied to the condition before or little research has been done in the area.	2	14
Increasing importance and recognition of preferences and quality of life issues in assessing outcomes of interventions	2	9
To facilitate economic evaluations of treatments		5
No justification provided	3	4
Due to inadequacy of existing instruments to measure quality of life issues	1	4
To contribute to resolving controversy over validity of use of utilities in CEA		4
Little is known about how preferences differ with rater characteristics		3
To facilitate comparisons		3
Little is known about differences in preferences between methods in different contexts		3
To study relationships between measures	2	2
Others		2

NB// totals do not add up to 42 because reasons were not mutually exclusive i.e. some studies provided more than one reason.

4.3.3 Methodological and empirical issues in assessing disease specific utilities

4.3.3.1 Measurement issues

Only 18 out of 42 studies incorporated measurement instruments and these were as diverse as the conditions studied. Table 4.4 shows the range of instruments by disease condition studied. No single instrument dominated except that SF-36 and HUI were used alongside other instruments in 7 and 3 instances respectively. Sixty one percent (11 /18) of the studies used a combination of disease specific and generic measurement instruments. Three studies used either disease specific or generic instrument separately.

A range of reasons was provided for the choice of measurement instruments, although 10 studies did not provide a reason. Choices were based on reliability, validity, responsiveness and wide usage (n= 9); ease of use and less demanding (n= 2); and 'other' (n= 6), including the monitoring of patients with multiple conditions, comparisons with general population, that published weights exist and could therefore be used to compute utilities, and that the measure had not been previously applied to the disease. The 'other' reasons for using disease specific instruments concerned evaluation of patients' ability to function or that the measure focused on the clinical condition.

Table 4.4: Measurement Instruments and conditions they were used in

Diseases or conditions	Disease specific instruments	Generic instruments
Angina	Index of co-existent disease (ICED) [12] 1 question for anginal severity [12]	SF-36[12]
Asthma	Asthma TyPE (Technology of patient experience)[33]	SF-36[33]
Breast cancer (metastatic)	EORTC QLQ-C30 [38] Breast reduction surgery attitude Questionnaire [23]	SF-36[23] EQ-5D[23]
Cataract / Visual loss	VF-14 [3,9] Snellen Visual Acuity [3, 9]	SF-36[9]
Cronic Airflow Limitation (CAL)	OCD visual analog score (1-10) [18] CRQ scores 0n each of 4 domains (1-7) [18] TDI index [18] Global rating of change on a seven-point scale [18] 6-minute walk test [18]	SIP[18]
Diabetes Mellitus	Clinical parameters of diabetes [5] Severity of co-morbidities associated with diabetes each assessed on a scale of 1-3 (4). [5]	
Depression	Hamilton Depression Rating Scale (HDRS) [4] Montgomery-Asberg Depression Rating Scale [1] Center for Epidemiologic Studies Depression Scale. [1] McSad (6 dimension s of depression; emotion, self-appraisal, cognition, physiology, behaviour, role function) [1]	
Fibromyalgia (FM)	Pain (VAS) [17] Global health (numerical rating scale) [17] Stiffness[17] AIMS[17] mHAQ [17]	MUMQ (Maastricht Utility Measurement Questionnaire) [7] SIP [17]
Heart failure	Minnesota living with Heart Failure questionnaire [16] 6-minute walk distance [16] New York Heart Association (NYHA) classification of heart failure (defines 4 states) [29]	SF-36[16]
HIV infection	MOS-HIV (adapted from SF-20) [37]	HUI[37]
Intermittent claudication		RAND 36 item survey 1.0[13] HUI 1[13]
Osteoporosis	Deyo Charlson Comorbidity Index [27]	SF-36[2] HUI 2[27] QWB[27]
Primary health care patients	Single item rating overall health[2] Count of number of chronic medical conditions[2] WHO 12-month composite International diagnostic instrument [2]	SF-12[2]
Rheumatoid arthritis	Clinical measures [28] Functional (HAQ, Keitel Assessement, Toronto Activities of Daily Living) [28] Pain (McGill Pain Q., Pain Ladder scale, 10 cm pain line) [21] Global impression [28] Arthritis (categorical scale and Ladder scale) [28] Overall health (ladder scale current, ladder scale 6-day mean. [28] Rand current health Assessment[28], 10 cm line patient[28], 10 cm line physician[28]) NIMH depression Questionnaire[28], Rand general health perceptions Q[28]. Erythrocyte sedimentation rate[28].	

1=Bennett et al. 2000, 2=Sherbourne et al. 1999, 3=Brown, 1999, 4=Ravichi and Wood, 1998, 5=Brown et al. 2000, 7=Kuijer-van Molken et al. 1995, 9=Lee et al. 2001, 12=Chen et al. 1996, 13=Bosch and Hunink, 1996, 16=Havranek et al. 1999,17=Bakker et al. 1995, 18=Guyatt et al. 1999, 23=Kerrigan et al. 2000, 27=Gabriel et al. 1999, 28=Bombardier et al. 1986, 29=Kirsch and McGuire, 2000, 33=Blumenscheinet al. 1998, 37=Bayoumi, and Redelmeier, 1999, 38=McLachlan et al. 1999.

Only two studies reported on the construct validity of the measurement tool. Sherbourne et al (1999) used the SF-12, single item rating of overall health, count of number of medical conditions and WHO-12 month composite International diagnostic instrument. Without saying how validity was assessed and on which instrument, they mention that validity of the items is supported by the findings that sicker patients had lower utilities for their current health. Guyatt et al (1999) compared the validity of a number of HRQL measures. This was tested by examining the a priori predictions about the magnitude of correlation the authors expected if the CRQ (Chronic Respiratory Questionnaire), OCD (Oxygen Cost Diagram) and SIP were measuring what was intended. Their findings showed that the CRQ and other disease specific measures were very responsive to change reaching high levels of statistical significance. SIP and OCD were not responsive to change and were not significant. None of the studies looked at the reliability or practicality of the measurement instruments used.

4.3.3.2 Valuation issues

4.3.3.2.1 Valuation techniques used

The most commonly used valuation instruments were TTO (n=26), VAS (n=23) and SG (n=21). PTO and ME were not used in any of the studies. TTO, SG and VAS were used alone 8, 5 and 4 times respectively. The most common combination was VAS and TTO (n=6) followed by TTO, VAS and SG (n=5). Generic preference based instruments like the QWB (n=2), EuroQol (n=2), and Maastricht Utility Measurement Questionnaire (MUMQ) (n=1) were also used to estimate preferences. The Patient Utility Measurement Scale (PUMS) which is a combination of variants of VAS, TTO and SG was used in one study.

Table 4.5 shows reasons provided for choice of valuation instruments. Although a variety of reasons were provided, most studies did not provide any reason.

Table 4.5: Reasons for choice of valuation instrument

Reason	RS	TTO	SG	WTP	PUMS
No reason provided	11	14	8	3	
Extensively used, popular and well established	4	4	5		
Has a theoretical foundation in VNM-EUT			7		
Preferences are measured under uncertainty			4		
Easier than SG		3			
Classic technique and original method for measuring utilities			2		
Very simple technique, easy to use	3	2	1		
Other	6	11	4	1	1

The most common reason provided for use of RS was that it was extensively used, popular and well established as well as being simple to use. Other reasons for choosing RS were that; it complements other methods, it avoids problems with comprehension and respondent burden; and that it was reliable. For the TTO, the arguments were the same except that three also stated that it was easier than the SG. Other reasons for the TTO were that: it was valid and reliable; it was more realistic and credible because it is choice based; it was not explored in the context it was used in before; for practical reasons on number of instruments that can be used; and due to lack of consensus on which to use. The common reason for choosing the SG was its theoretical foundation in von-Neuman-Morgenstern expected utility theory and use under conditions of uncertainty. However, other reasons also included an instrument being extensively used, choice based or simple.

4.3.3.2.1 Validity assessment

Eight studies explicitly, and three implicitly, investigated the validity of the valuation approach adopted. Construct validity was the most commonly assessed form of validity (n=9). Only two studies assessed either content or criterion validity. Construct validity was assessed more often for RS than TTO and SG, both of which were assessed equally (table 4.6). No studies assessed either construct or content validity of the SG.

Table 4.6: Validity assessment by type and instrument

	RS	TTO	SG
Content validity	1	2	
Construct validity	7	4	4
Criterion validity	1	2	

4.3.3.2.2 Content validity

Two studies assessed content validity for TTO in the UK and USA amongst breast cancer and rheumatoid arthritis patients respectively (Gerard et al. 1999 and Gabriel et al. 1994). Gabriel et al (1994) provided no explicit details of how this was done but simply declared face and content validity based on views of independent assessment by three experts. Gerard et al (1999) assessed descriptive validity of four breast screening outcome descriptions by asking subjects to map them onto the EuroQoL classification system to gather evidence on respondent's perceptions of the content of the scenarios. The frequency with which different EuroQoL states were selected to represent the disease specific descriptions was used to judge the descriptive validity of the scenarios based on the assumption that the EuroQoL is descriptively valid. The disease specific descriptions were deemed invalid based on the heterogeneity of perceptions in that no EuroQoL descriptions dominated in describing the disease scenarios.

4.3.3.2.3 Construct validity

Nine studies assessed construct validity. Six studies done in North America assessed construct validity for VAS (n=5), TTO (n=4) and SG (n=4) using varied conditions including dental implants, chronic airflow limitation, heart failure, rheumatoid arthritis, breast hypertrophy and prostate cancer. Two studies in Europe assessed construct validity for VAS, SG and TTO, each once amongst arthritis and breast cancer patients. In Africa, Baltussen et al (2002) assessed construct validity for an adapted version of VAS using nine different disease conditions (see appendix 4.4 for details). Five studies had samples ranging between 47 and 89 subjects while the rest ranged between 111 and 440 subjects. Seven studies had objectives directly related to evaluating the validity of the instruments.

Six studies stated hypothesized relationships expected in their construct validity testing, although two (Guyatt et al. 1999 and Gabriel et al. 1994) were not explicit. As appendix 4.4 shows, most of these hypotheses suggest that worse off health states would be assigned lower utility scores in cases where the hypothesis relates to assessing changes and or differences in health states. Similarly, where hypotheses relate to

different methods being correlated, the authors hypothesize that there would be significant relationships between utility scores and other HRQL measures (appendix 4.4).

The most commonly used analytical method for assessing construct validity was correlation analysis (n=7) mainly using Spearman's and Pearson's correlation coefficients (appendix 4.4). Four studies made comparisons of mean utility scores using a variety of statistics. Jacobson et al (1992) used a one-tailed *t-test* while Kerrigan et al (2000) used Wilcoxon rank-sum test. Souchek et al used a differentiation and inconsistency score[29] while Gabriel et al (1994) failed to provide the test they used. Only one study Bakker et al (1995) used multiple regression analysis with changes in utility scores as the dependent variables and treatment changes in other health outcomes as independent variables.

The majority (n=7) of studies tested construct validity by correlating utility scores with different measures of HRQL, both generic and disease-specific as well as correlating scores between different valuation techniques. For example, Havranek et al (1999), correlated TTO and VAS scores with 6-minute distance walk, SF-36 HRQL questionnaire results and a visual analogue score of overall health found significant relationships between utilities and these measures such that higher scores corresponded to less severe disease. In Baker et al (1995) construct validity of RS utilities was supported by significant correlations with measures of global health, pain, SIP, AIMS, mHAQ while SG had lower construct validity as it correlated considerably less with global health, SIP, AIMS, mHAQ, pain and stiffness. Other studies using similar methods include Guyatt et al (1999), Kerrigan et al (2000), and Gerard et al (1999). Souchek et al (2000) correlated utility values between pairs of RS, TTO and SG and reported moderate to high correlation (RS/SG-0.74; TTO/SG-0.69; RS/TTO- 0.76) thereby supporting construct validity of RS, TTO and SG. While Gabriel et al (1994) suggest convergent validity for TTO and VAS, the authors did not state how it was assessed.

[29] Using ranking as a gold standard, a differentiation index for a technique indicates how often outcomes should have received (according to ranking) different utility values actually did receive different utility values and the inconsistency index indicates how frequently utilities indicated a preference ordering opposite to preference ordering expressed in the ranking exercise. (See Geisler et al, 1999 for details of computation of the two indices).

Four studies assessed construct validity by hypothesizing and examining how utility values changed and or differed for different health states. For instance, Jacobson et al (1992) hypothesized and found that dental implants pre-insertion mean score was significantly lower than post-insertion score thereby supporting RS construct validity. Gabriel et al (1994) using VAS and TTO and Kerrigan et al (2000) using VAS and SG demonstrated that worse off health states received lower values. In Gabriel et al (1994) utility score for 'no ulcer' was significantly higher than 'ulcer requiring medical treatment' which was significantly higher than that for 'ulcer requiring surgery' and 'Prophylaxis without side effects' was scored higher than 'prophylaxis with side effects'. Similarly in Kerrigan et al (2000), women with breast hypertrophy had lower utilities for current health than those without while those rating their health (SF-36) as excellent or very good had higher utilities than those rating it as good, fair or poor. Bakker et al (1995) examined discriminant validity of SG and VAS and found significant correlations in changes in RS scores and changes in four dimensions of AIMS, SIP, pain and global health while in SG, significant correlations were with some dimensions of AIMS, and global health. Using regression analysis, changes in RS were explained to a higher degree than SG, supporting better discriminant validity of RS than SG.

Baltussen et al (2002) assessed the construct validity of an adapted version of VAS by correlating values from lay people and health professionals. They found high (0.86 and 0.94) correlation coefficients supporting construct validity of VAS.

It is worth noting that no valuation technique was declared invalid as all studies reported validity to varying degrees. However, with the exception of Souchek et al (2000), no study stated an a priori level of confirmation for hypothesized relationships or correlation coefficients at which methods would be declared valid or invalid.

4.3.3.2.4 Criterion validity

Only two studies (Swan et al. 2000 and Mackeigan et al. 1999) assessed criterion validity. In Mackeigan et al (1999), rating scale ranking of scenarios was used as the validity criteria for TTO based on the argument that VAS produced a direct rank

ordering. Mean rankings from the TTO valuation exercise were compared with the VAS direct ranks. They reported poor agreement between rank order preferences implied by the TTO based composite and holistic preferences and the VAS direct preference rankings, thereby casting doubt of the criterion validity of TTO. In Swan et al (2000), it is not clear how criterion validity was assessed since they only asked respondents to do a rating on a direct rating scale. There is no report on its findings. Rather, the authors refer to face validity (another type of validity) in the discussion without noting if and how it was assessed.

4.3.3.2.2 Reliability assessment

Ten studies assessed test-retest reliability, making it the most commonly assessed form of reliability, while one study assessed inter-rater reliability and one internal consistency. Table 4.7 presents information on assessment and empirical findings on test retest reliability.

The period between test and re-test ranged from 1 day to 6 months, with most studies having an interval of 3-6 weeks. The sample sizes for retest ranged from 7-228 subjects with the majority having 50 or less respondents (n=6). Commonly used statistics for test-retest reliability were the Spearman correlation coefficients, interclass correlation coefficients and testing for dispersion of values using variance. Other statistics used less commonly were the Pearson correlation coefficient, the *t-test* statistic for differences between mean scores, and standard errors to show variability in values.

Interclass correlation coefficients for RS ranged from low (0.24-0.33) to high (0.713 and 0.82) while Spearman's rank correlation coefficients reported for RS was 0.47. Other measures show that RS is reliable. TTO shows moderate to high Spearman's correlation coefficients (0.5-0.79) and no significant differences in values between test and retest. However, in terms of variability, TTO shows instability and wide variation in values ranging form 0.21-1.0. Interclass correlation coefficients for SG range from 0.43-0.7 while Spearman's correlation coefficient is 0.59.

Like the TTO, the test of stability showed poor and unstable results for SG. Therefore, different test statistics give conflicting results regarding reliability of TTO and SG.

However, judging by the results of the commonly used test statistics, the two methods seem to have moderate to high test-retest reliability. As with validity testing, there is no set critical correlation coefficient levels below which methods are deemed unreliable. Hence, conclusions about reliability appears to depend on what level the authors decide.

Table 4.7: Assessment of test-retest reliability (n=10)

Duration between test and re-test	Retest sample size	Test statistic	Empirical findings		
			RS	TTO	SG
6 months	20[6]	Inter class correlation coefficient	0.713		
	73[17]	Stability of marker states	Poor and unstable		Poor and unstable
3 months	Not stated[7]	Interclass correlation coefficients	0.24-0.33		0.43-0.7
		Standard errors of values	9		0.12
3-6 weeks	50[11]	Spearman correlation coefficient		0.5-0.75 (all significant[a]	
	7[19]	Pearson correlation coefficient	0.89	0.63	
		Variability of scores	Between 0.81-0.96 (Excellent)	Between 0.21-1 (variable)	
	56[25]	Pearson correlation coefficient	0.90 (lay people) 0.89 (health Profs.)		
	44[26]	Stability of values		Poor test retest	
1-3 weeks	12[16]	Differences between values		No significant difference between baseline and retest values (0.77 and 0.78)	
	47[23]	Interclass correlation coefficient	0.82		0.60
		t- test	No significant difference		No significant difference
		Spearman rank correlation coefficient	0.47		0.59
1 day (not same sample)	228[2]	Spearman correlation coefficient		0.79	0.69

a significance levels refer to 95% level of confidence and above

2=Sherbourne et al. 1999; 6=Jacobson et al. 1992; 7= Rutten-van Molken et al. 1995; 11= Ashby et al. 1994; 16= Havranek et al. 1999; 17= Bakker et al. 1995; 19= Gabriel et al. 1994; 23= Kerrigan et al. 2000; 25= Baltussen et al. 2002; 26= Gerard et al. 1999.

4.3.3.2.3 Assessment of practicality

Seven studies assessed practicality. RS and TTO were assessed four times each and SG was assessed twice. Time taken to complete the interview was the most commonly used indicator of practicality (n=4) followed by comprehension of the task as judged by the interviewers (n=3) and ease of use as rated by subjects on a scale of very easy to very difficult (n=3). Other indicators of practicality include ability to effectively answer questions, acceptability of the task, and extent of incomplete date.

In Bakker et al (1995), it took 10.8 (SD 2.9) minutes at baseline and 9.4 (SD 7.3) minutes six months later to value four health states using the RS while Gabriel et al (1994) recorded 42 minutes (no SD) to value eight states. In Burkina Faso, it took 74 minutes to describe nine hypothetical states and explain the instrument, 52 minutes to arrive at individual valuations and 12 minutes to make final assessments (no SD) (Baltussen et al. 2002). Ashby et al (1994), valued six states and reported that TTO is relatively brief and easy to administer (no time provided) while Gabriel et al (1994) used 40 minutes (no SD) to value eight states. In Baker et al (1995) it took 12.5 (SD 3.8) minutes to value four states using the SG at baseline and 11.5 (SD 4.5) minutes six months later. Even though the studies have assessed the time it takes to complete the valuation task, there is no uniformity in terms of the aspects assessed. For example, it is not clear whether these timings related to valuation only or whether they include other aspects of the interview. Also the different valuation techniques are not assessed for time it takes to complete using the same standard (e.g. time from start to end of each valuation task within the overall interview). Therefore at this point, we cannot conclusively say which valuation task takes less time than the other does. Also, despite the fact that different studies valued different number of states, it was not straightforward to make comparisons between time taken and number of scenarios due to the small number of studies assessing time and lack of information on what aspect of the task was valued.

In terms of comprehension of the TTO, Gerard et al (1999) reported that 93% of the respondents understood what was being asked as judged by the interviewers, while

Swan et al (2000) deemed it satisfactory without explaining how this was assessed. Bakker et al (1995) reported that very few respondents gave inconsistent answers for the SG and RS without explicitly reporting on how this inconsistency was defined and assessed. None of the studies assessed ease of use of the SG. Based on a rating scale of 0 (very easy) to 100 (very difficult), people rated the TTO at 40 compared with the RS at 30 (Gabriel et al. 1994). Gerard et al, found that 84% of the respondents found TTO very easy to fairly easy.

Other aspects of practicality assessed revealed that RS was well accepted based on evidence that no interviews were broken off. For TTO, 93% of respondents were able to effectively answer questions as judged by the authors (Brown, 1999), it was found acceptable (not explicit how this was assessed) (Ashby et al. 1994) and only 2.7 % of data was incomplete in Gerard et al (1999). For SG, 90.4% of respondents were effectively able to answer questions (Brown, 1999) and it was acceptable judged by the fact that no interviews were broken off (Bakker et al. 1995).

4.3.4 Other methodological and empirical issues in measurement and valuation of disease utilities

4.3.4.1 Cross cultural issues

Cross-cultural issues mainly relate to use of instruments in settings in which they were not developed. Considering that VAS, TTO and SG were all developed in North America, we can assume that the 31 studies undertaken there did not require cross-cultural adaptation[30]. Out of eight studies undertaken in the UK and the Netherlands, only Ashby et al, 1994 assessed cross-cultural issues. Hence, out of 11 studies used in other settings, only 3 acknowledged that the valuation instruments were being applied in a different setting (Ashby et el. 1994, Kirigia, 1998 and Baltussen et al. 2002). Two of these involved use of the RS (VAS) and one the TTO. The VAS studies were both in Africa and involved modifying the VAS to make it easier and more culturally relevant

[30] It can also be argued that this assumption is unrealistic. On an absolutist-universalistic-relativist approach as described in Herdman et al. 1997, it takes an absolutist view to culture. Such an assumption wishes away the influence of culture on different aspects of health of indigenous peoples in America, Canada and New Zealand as discussed in an editorial in British Medical Journal volume 327, August 2003. Hence, the 31 studies ought to report on cross-cultural issues depending on the populations they are used on.

to the new populations. In Kenya, Kirigia (1994) modified the VAS to a rice-sack VAS (as he worked in a rice farming community) while in Burkina Faso, Baltussen et al (2002) used physical units (10 6-cms wooden blocks) rather than numbers to represent the VAS. Both studies based their justification for the modification on low levels of education in these populations and in Burkina Faso, response spreading in using traditional VAS. Ashby et al (1994) used TTO in the UK without modifying it and then compared its performance with Canadian studies while Baltussen et al (2002) pre-tested the VAS in its original form before deciding to adapt it for the rural Burkina Faso population. Kirigia (1994) neither pre-tested nor compared the performance of the VAS in the Kenyan setting before deciding to adapt it.

With respect to measurement instruments, two studies (Bakker et al. 1995 and Rutten van-Molken et al. 1995) acknowledge using an adapted version of the HUI or the Maastricht Utility Measurement Questionnaire (MUMQ). Although they mention that the instrument was adapted, no details were provided of the adaptation process, making it difficult to comment on the suitability of the measures in their new settings or for the specific conditions to which they were applied.

4.3.4.2 Elicitation of values/utilities

All studies elicited preferences either as utilities, values or both. Twenty studies assessed values (i.e. used TTO or RS), 7 studies elicited utilities (i.e. used SG) and 15 elicited both values and utilities. Of the 42 studies only 14 measured disease specific outcomes while 13 included other HRQL measures. Ten studies assessed both disease specific and other HRQL outcomes in addition to preferences. This suggests that the majority of studies elicited only preferences without relating them explicitly to measurements of HRQL.

In terms of number of scenarios valued, the majority of studies (n=24) valued between 1-4 scenarios while another 15 valued 5-9 scenarios. Three studies valued 12, 24 and 25 scenarios each. Table 4.8 shows the type of scenarios and whether they were measured during the study or not, indicating the extent to which preferences elicited were or were not linked to measurement of health states from the subjects. Twenty and eight studies

used hypothetical[31] and treatment option as scenarios respectively and 16 valued 'current health-state'.

Table 4.8: Type of scenario valued and whether measured or not

	Hypothetical scenario	Current health state**	Treatment options and diagnostic tests
Measured from subjects doing valuation*	1	16	1
Not measured from subjects doing valuation	19	2	7

* Note that entries in this row refer to implied measurement rather than explicit measurement of HRQL.

** Entries in this column refers to studies valuing current health and administering other instruments at the same time. No explicit links are made between what is valued and measured although it is implied.

One study (Douzdjian et al. 1998) used hypothetical states which the patients were assumed to have experienced earlier on and consequently no explicit descriptions of these states were provided to the patients. There was thus some implied link to earlier measurement. Sixteen studies valued 'current health-state' or 'your present state of health' or 'own health-state'. In nearly all of these studies, other HRQL instruments were administered to the respondents during the interview, but it is not stated whether 'your current health' was described according to the instruments' descriptive system. For example in Bombardier et al (1986), about 20 instruments were administered at differing points during the trial and subjects were asked to value their current health state and recollected health state prior to the trial. From this study, one cannot tell what 'current health state' consists of (in terms of descriptive information) and which instrument the authors and or the subjects used to define and describe their current health state. None of the studies using 'current health-state' as a scenario provided explicit information linking the components of the scenario to a measurement instrument. It was therefore left to the rater to provide their own description and interpretation of the scenario they were asked to rate. Although by labeling, the subjects may seem to have valued the same state, we cannot tell what each subject considered and rated, which would make it extremely difficult to compare valuations from these studies.

[31] Hypothetical scenarios refer to those states that were not directly obtained from subjects doing the valuation although they could have been related to the conditions the subjects were having. For example, Stolk et al, 2000 used both men and women to value 25 erectile dysfunction scenarios based on International Index of Erectile Dysfunction. Treatment options are for example in Swan et al (1997) who valued two angiography tests (MRA and XRA) or in Hayman et al (1997) who valued 5 treatment options for breast cancer with combinations of radiation therapy, conservative surgery, mastectomy and reconstructive surgery, expressed with various levels of risk of local recurrence.

Four studies valued 'current health-state' alongside some hypothetical states while 15 studies rated purely hypothetical states. In total, 19 studies valued hypothetical states that were not related in any way to health states measured from the patient population or other raters doing the valuations, either previously or during the study.

Seven studies valued either treatment options or diagnostic tests (see footnote 32). For example MacKeigan et al's (1999) eight-treatment paths comprised three oral mono therapy states, three dual therapy, triple oral therapy and insulin for diabetes type 2. Each treatment path contained information on frequency of drug administration, blood glucose testing, drug efficacy and frequency and nature of side effects. An interesting observation was that these studies referred to these as health states, although they were not described in any way relating to HRQL. There seems to be confusion in the usage of the term health-state and what is being valued in the literature.

4.3.4.3 Content of scenarios and their justification

Thirty-two studies provided descriptions of scenarios that were valued. Table 4.9 classifies these descriptions into 7 main categories; functioning, mental well being, physical appearance, social participation, symptoms, disease labels and treatment options. Under functioning common descriptors include ability to look after one-self or self-care functions, ability to work, perform activities of daily living, role functioning and mobility. Common descriptors classified under mental well being include anxiety, depression, emotions, frustrations, worries and mood. Other less used descriptors are behavior, cognition and well being. Physical appearance and social participation are used less often. About 21 studies used symptoms as part of their scenario descriptions. The most commonly used symptoms are pain and discomfort. Other symptoms are specific to the conditions and treatment side-effects. Five studies used disease names to label and refer to their scenarios without necessarily giving any explicit description of the scenario. Six studies elicited values for treatment options, diagnostic tests, hospitalization and therapy.

Table 4.9: Descriptors used in scenarios

Broad domain	Times descriptor used	Descriptors	No. of studies
Functioning	7	ability to look after oneself / personal care (self care functions); [17, 37]	13
	2	[27, 29, 30,7, 8]	
	7	ability to work / livelihood [8, 37]	
	5	activities of daily living /usual activities [7 ,17,25,27,29, 30, 36]	
	5	functioning /role function [1,10, 6, 4, 29]	
		mobility [23, 27, 29, 30, 8]	
Mental well being	10	anxiety/ depression/ emotions, frustrations and worries, Mood	11
	1	[5,37,1,29,30,7, 17,11, 10, 27]	
	1		
	1	behavior [1]	
		cognition, [1]	
		wellbeing, [4]	
Physical appearance	3	physical appearance [6, 11,1]	3
	1	self appraisal [1]	
Social participation	3	social participation, [11, 8, 10]	5
	2	leisure activities; [17, 7]	
Symptoms	1	clinical aspects of ulcer and complications [19,	21
	1	diarrhea/nausea [40]	
	2	energy level, [36, 8]	
	1	fatigue / tiredness, [36, 40]	
		feelings of sexual desire /loss of libido /ability to achieve an	
	3	erection, [32, 36, 40]	
	2	hot flushes, [36,40]	
	1	loss of body hair [36]	
	1	muscle tone [36]	
	10	pain and discomfort [27,10,7 , 15, 36, 40, 8,29, 30, 17, 6,	
	8	side effects [31,36 17, 19, 21,42, 4,7]	
	8	symptoms [19, 21, 22, 4,29, 8, 37, 36]	
Disease labels	1	any medical conditions you have' [27]	5
	3	disease names (labeling by disease names) [25, 24, 27]	
	1	progression of disease[41]	
	1	psoriasis (% of body covered) [22]	
	1	response [41]	
Treatment options	1	frequency of clinical tests, [21]	6
	1	frequency of physician visits [36]	
	1	hospitalization episodes, [37]	
	1	medication therapy, [4]	
	3	treatment choice [36, 31, 41]	

1=Bennett et al. 2000; 4=Revicki and Wood, 1998; 5=Brown et al. 2000; 6=Jacobson et al. 1992; 7= Rutten-van Molken et al. 1995; 8=Kirigia, 1998; 10=Johnston et al. 1998; 11= Ashby et al. 1994; 15=Swan, et al. 1997; 17= Bakker et al. 1995; 19= Gabriel et al. 1994; 21=Mackeigan et al. 1999; 22=Chen et al. 1998; 24=Douzdjian et al. 1998; 25= Baltussen et al. 2002; 27=Gabriel et al. 1999; 29=Kirsch and McGuire, 2000; 30=Sauraez-Almazor et al. 2001; 31=Hayman et al. 1997; 32=Stolk et al. 2000; 36=Souchek et al. 2000; 37=Bayoumi, and Redelmeier, 1999; 40=Matchar et al. 1997; 41=Dranitsaris et al. 2000; 42=Leung et al. 1990.

Sixteen studies did not provide any information on how the scenarios were determined and justified. Of the 26 providing information, 20 did not justify the methods they used to determine the scenarios to value while only 6 did.

The most common method used to determine the descriptive basis of scenarios was to use an existing health or disease state classification system such as McSad (full name not given), EQ-5D, VF-14 (Visual Function 14), MUMQ (Maastricht Utility Measurement Questionnaire), NYHA (New York Heart Association) heart disease classification and international index of erectile dysfunction. Both disease specific and generic instruments were used for this purpose. Other methods for determining scenarios were reviews of literature (n=5), consultation with experts in the field (n=4), expert judgement (n=3) and focus group discussions either with people with the condition, have experience with the condition and physicians (n=3). Less used methods included; asking patients to recall their experience (n=2), authors providing the scenarios (n=1), based on available treatment options (n=1), guidelines (n=1) and using scenarios from other studies (n=1).

For those studies providing justification for how they determined their scenarios (n=6), the reasons were varied. Johnston et al (1998) used a pragmatic approach to describe their scenarios because there was no consensus in the literature on the dimensions affected by breast screening. Ashby et al (1994) on the other hand chose their scenarios to reflect consideration of both the number of different scenarios that could be reasonably inferred from literature and the number that could be reasonably presented in an interview. The other four studies (Swan et al. 1997; Douzdjian et al. 1998; Baltussen et al. 2002 and Hayman et al. 1997) presented scenarios relating to treatment options and diagnostic tests and therefore were guided by what treatment options and therapies they were evaluating.

4.3.4.4 Whose values and why

Table 4.10 presents the type of raters chosen and valuation techniques used. Thirty-three studies elicited preferences from patients, 12 from general population (farmers, teachers, university staff, and lay people) and 11 from health professionals (nurses, general practitioners, physicians and hospital doctors). For the TTO and RS, patients were chosen nearly five times more often than health professionals and 2.5 times more times than the general population. For the SG, patients were chosen four times more than the health professional and 10 times more than the general population. The SG was used more with the health professionals compared with the general population. The

TTO and RS were used to the same extent with both the general population and the health professionals, although less often with the health professionals.

Table 4.10: Choice of raters by valuation techniques

	RS	*TTO*	*SG*	*Number of times rater chosen* [a]
Patients	*20*	*19*	*20*	*33*
General population	*8*	*8*	*2*	*12*
Health professionals	*4*	*4*	*5*	*11*
Number of times technique used [a]	*23*	*26*	*21*	

[a] totals do not add up to 42 because both raters and valuation techniques were not mutually exclusive. However the totals indicate extent of use of both the technique and the rater.

Most of the studies using patients to elicit preferences (16 of 33) did not provide any justification for choosing patients as the raters. The most common reason for using patients when given, was that they are the best source of information about their own health and quality of life because they have first hand experience of the condition and only they know the true implications of a particular health state. Alternatively, researchers wanted to compare preferences between different groups of raters to account for the controversy of whose values to use in measuring utilities (see table 4.11). Other reasons included the fact that patients are the beneficiaries of services provided and should therefore provide values. In four out of 11 studies health professionals were chosen to provide values for comparisons to aid decisions on whether health professionals' preferences could be used as proxy for those of the community and or patients. Two studies using general public as sources of preferences provided no justification for their choice. Four studies justified their choice by arguing that their study was taking a societal view while another four justified their choice as providing preferences for comparisons between groups of raters to inform on whose values should be used.

Table 4.11: Justification for choice of raters (No. of studies)

Justification	Patients	Health profs.	General population
No reason provided	16		2
Preferences (utilities) not elicited from this patient group before	3		
Differences in preferences between patients and other groups	3		
Patients are the best source of information about their health because they have experienced the condition	8		
To eliminate the problem of having to describe health states to general public	1		
Appealing to use patients	1		
To reflect preferences of those who should benefit from the services provided	2		
Groups at risk of infection	1	1	1
For comparisons between groups to inform on the issue of whose values should be used	6	4	4
To adopt a social viewpoint			4
To act as proxy for patients because they are familiar with the health states		3	
To exclude bias from patients		1	
Their judgements strongly influences therapy selected		1	
To follow guidelines (Canadian and Gold et al, 1996)		1	1

4.3.4.5 Factors affecting values for the same disease state

Fifteen studies assessed factors affecting values, eight of which had sample sizes ranging between 50-100, five had 101-500, one over 500 and another 17,107 subjects, reflecting fairly moderate to large samples. Several variables were reported in the reviewed studies. They include age (n=13), gender (n=10), education (n=9), marital status (n=5), duration of state (n=5), ethnicity (n=3), severity of state (n=3), presence of co-morbidities (n=3), experience with disease (n=3), whether sick or not (n=2), income (n=2), household size (n=1) and working status (n=1). Age, gender and education were the most commonly reported variables.

Considering the number of studies assessing each factor, in the majority of cases no statistically significant relationships were found between values and or utilities and the factors. There was no significant relationship at all between household size, working status and income (socio-economic status). The few that found significant relationships had mixed findings as reported in table 4.12. Amongst them, except for gender (n=3), age (n=2) and experience with disease (n=2), other variables with mixed findings had only one study in each case reporting significant relationships (table 4.12), in most cases

by Sherbourne et al (1999)[32]. In general significant negative relationships were found between values/utilities and age (Sherbourne et al. 1999), education level (Sherbourne et al. 1999) and co-morbidity (Bosch and Hunnik, 1996). Also in Sherbourne et al (1999), men, the unmarried and the whites (in relation to the blacks) had lower values. In situations where positive and significant relationships were found, only experience with disease (Ashby et al. 1994; Gabriel, et al. 1999) and whether sick or not (Sherbourne et al. 1999) had consistent findings that patients had higher values. There was also a positive relationship between values and age (Kirsch and McGuire, 2000), while men (Bosch and Hunnik, 1996; Kirsch and McGuire, 2000) and the whites (in relation to the Asians and Latinos) (Sherbourne et al. 1999) had higher values.

Duration of state and severity of state affect values for different disease states. Out of five and three studies that examined the factors respectively, only one in each case found significant relationships. Kirsch and McGuire (2000) found higher values for the most severe disease state with shorter duration than one with longer, while Revicki and Wood (1998) found lower utilities for major depression hypothetical disease states from those with most severe depression.

With the exception of Sherbourne et al (1999), no other study gave reasons and or justified why certain factors were evaluated and not others. Sherbourne et al noted that little was known about how characteristics of the rater influence value judgements and whether patients place a higher value on quality of life than length of life. In general, none of the studies reviewed presented the hypothesized relationships expected between values and the variables or how such hypothesized relationships derived from theory.

Three main techniques were used to assess the relationships between variables and values. Comparisons of mean and median values between pairs and groups (ANOVA) (n=11) and regression analysis (n=10) were the most commonly used methods, with a fewer (n=5) studies using correlation analysis. Studies using the technique of comparisons of mean values tended to use a mixture of both parametric and non-parametric statistics. Parametric tests included the student *t-statistic,* the *F-statistic*, the chi-squared statistic as well as presentation of mean values and their standard

[32] This study had an exceptionally large sample size: 17,707 patients.

deviations. Non-parametric statistics included the Wilcoxon matched pairs, Mann-Whitney *U*, Kendall's coefficient and Kruskal Wallis statistics. Studies using regression methods utilized the *beta* coefficients, correlation coefficients, *t-statistic, F-statistic* and the R^2. The majority (n=6) of studies using regression analysis failed to provide their test statistic. Studies using correlation analysis reported both Spearman's and Pearson's correlation coefficients. There was a lot of variability in methods and test statistics used.

Table 4.12: Summary of extent of exploration, analytical tools, and relationships between factors and values/utilities

Variable	Number of studies considering variable	Analytical tools and test statistics used	Number of studies finding statistically significant relationship between variable and values/utilities	Direction of relationship between variable and values/utilities
Age	13 [1,2,3,4,6,7,8,10, 11, 12,13,14,15]	• Regression: R^2 [1,10,12, 14,15,] • Comparison of means and medians: t-statistic [2,3,4,6,13]; Mann-Whitney [6]; Wilcoxon matched pairs statistic [6,7]; Kendall's coefficient [7] • Correlation analysis: Pearson's [8,14]; Spearman's [11]	2 [1,13]	• Negative: lower utility observed for the oldest patients [1] • None: No difference in mean values by age groups [2, 3, 4, 6,7,8,10,11,12,14,15] • Positive: In worse off disease states, the elderly had higher values [13]
Gender	10 [1,2,3,4,5, 7,9,10, 13,14]	• Regression: betas [1,5,14]; F-statistic [10] • Comparison of means and medians: t-statistic [2,3,4,7,9,13] ; Mann-Whitney statistic [9] • Correlation analysis: Pearson's [14] Spearman's [10]	3[1,9,13]	• Men had lower values than women [1] • None: [2,3,4,5,7,10,14] • Women had lower scores than men [9, 13]
Education	9 [1,2,3,4,5,8,11, 12,14]	• Regression: betas [1,5,12] • Comparisons of means and medians: t-statistic [2,3,4] • Correlation analysis: Pearson's [8,14] Spearman's [11]	1[1]	• Negative: The more highly educated had lower values and utilities [1] • None: No difference in values by education [2,3,4,5,8,11,12,14]
Marital status	5 [1,3,5,8,11]	• Regression: betas [1,5] • Comparisons of means and medians: t-statistic [3,8] • Correlation analysis: Spearman's [11]	1[1]	• Unmarried respondents had lower values [1] • No relationship [3,5,8,11]
Duration of state	5 [2,4,8,11,13]	• Comparisons of means and medians: t-statistic [2,4]; Wilcoxon matched pairs statistic [13] Kruskal Wallis [8] • Correlation analysis: Spearman's [11]	1 [13]	• For the most severe state, that with shorted duration receives higher value than one with longer duration. [13] • No relationship [2,4,11]

Ethnicity	3 [1,2,4]	• Regression: betas[1] • Comparisons of means and medians: t-statistic [2,4]	1[1]	• Asians and Latinos had lower utilities than Whites, whereas Blacks had higher utility than whites[1] • No relationship[2,4]
Severity of state	3 [3,14,15]	• Regression: betas [15] • Comparisons of means and medians: t-statistic [3] • Correlation analysis: Pearson's [14]	1 [3]	• Those with more severe depression provided lower utilities[3] • No relationship [14,15]
Presence of co-morbidity	3 [4,9,12]	• Regression: [12] • Comparisons of means and medians: t-statistic [4,9]	1[9]	• Patients with co-morbidity had lower values than with only one disease. [9] • No relationship [4,12]
Experience with disease	3 [5,7,12]	• Regression: betas[5] • Comparisons of means and medians: Wilcoxon matched pairs statistic [7,12]	2[7,12]	• Patients' values higher than for other subjects [7,12] • No relationship [5]
Whether sick of not	2 [1, 8]	• Regression: betas[1] • Comparisons of means and medians: Wilcoxon matched pairs statistic [8]	1[1]	• Utility was higher in patients [1] • No relationship [8]
Income (socio-economic status)	2 [10,12]	• Regression: F-statistic[10]; R^2 [12]		No relationship [10,12]
Household size	1[5]	• Regression: Betas[5]		No relationship [5]
Working status	1[8]	• Comparisons of means and medians: Wilcoxon matched pairs statistic [8]		No relationship [8]

1=Sherbourne et al. 1999; 2= Brown, 1999; 3= Revicki and Wood, 1998; 4= Brown, et al. 2000; 5= Kirigia, 1998; 6= Johnston et al. 1998; 7= Ashby et al. 1994; 8= Chen et al. 1996; 9=Bosch and Hunnik, 1996; 10=Havranek et al. 1999; 11= Bakker et al. 1995; 12=Gabriel, et al. 1999; 13=Kirsch and McGuire, 2000; 14= Zug et al. 1995; 15=Bayoumi and Redelmeier, 1999.

4.3.4.6 Intended use of disease specific utilities

Suggested uses for disease specific utilities (DSU) can be categorized into three groups namely; economists[33] concerned with resource allocation decisions using cost utility analysis (CUA); clinicians[34] concerned with patient management decisions and; patients[35] concerned with their health related quality of life. The most common use of DSU was in CUA. In this connection, the most (n=15) suggested use was as quality weights in constructing QALYs, (Lee et al. 2001; Stolk et al. 2000; Brazier and Dixon, 1995; Drummond et al. 1997; Revicki and Wood, 1998; Brown et al. 2000; Rutten van-Molken et al. 1995, Lenert et al. 1999; Swan et al. 1997 and Chen et al. 1998), to guide decisions about the economic value of therapy (Yee, 1997 and Douzdjian et al. 1998) and as parameters in decision models (Brown et al. 2000; Chen et al. 1998 and Douzdjian et al. 1998). Hence for economists, DSU are proposed as decision aids in making choices between alternative ways of allocating limited resources among different health care activities servicing the same or different patient groups or populations (Rutten van-Molken et al. 1995; Swan et al. 2000; Leung et al. 1999 and Gabriel et al. 1999).

DSU were considered important in aiding decision-making about choice of treatment options and patient management guidelines, whether individuals or groups (Bombardier et al. 1986; Hall et al. 1992; Revicki and Wood, 1998; Zug et al. 1995; Sherbourne et al. 1999; Gabriel et al. 1994; Mackeigan et al. 1999; Hayman et al. 1997; Souchek et al. 2000; Oldridge et al. 1993 and Zug et al.1995). DSU were also proposed as useful in indicating clinically important changes to patients (Zug et al. 1995; Brown, 1997; Lenert et al. 1999; Swan et al. 1997 and Chen et al. 1998).

DSU are also seen as a way of incorporating quality of life issues in health status measurement. For instance in Bombardier et al (1986) and Brown et al (2000) DSU were

[33] This term is used broadly here to include all concerned with issues of resource allocation in health care.
[34] This term is also used to broadly include all those involved in decisions concerning treatment choices and patient management.
[35] This is used to broadly include not only patients but the general population and health professionals who may have interests in health related quality of life issues of the population.

used to measure changes in patients' quality of life following treatment, thereby allowing assessment of the quality of life associated with a disease state (Brown et al. 2000). In a similar use, Brown (1999) noted that a drop of visual acuity from 20/70 to 20/100 (these clinical measures indicate that a patient is able to perform some activities and not others) was associated with a major drop in utility points, thereby indicating that these changes were valued more by the patients. There are also views that utility is a generic quality of life measure that theoretically allows a more comprehensive assessment of quality of life since it incorporates all parameters that comprise quality of life (Rutten van-Molken et al. 1995 and Lee et al. 2001; Brown et al. 2000). However, this may not be the case in practice.

4.4 DISCUSSION

Assessment of disease specific utilities (DSU) is just beginning to gain popularity. The bulk of research has been carried out in North America and Europe amongst chronic diseases. There is virtually no research on parasitic diseases such as malaria and schistosomiasis that afflict majorities of population in low-income countries. DSU are therefore a neglected area of research both spatially, and in terms of conditions prevalent in low-income countries like Kenya.

The motive and justification for undertaking the reviewed studies was mainly driven by the need to facilitate economic evaluation in the form of utility analysis and to a lesser extent help assess the suitability of the techniques. In addition, the increasing importance and recognition of preferences and quality of life issues in assessing outcomes of interventions was used as a justification for undertaking several studies. Johannesson et al (1996) envisages five main areas where economic evaluations are potentially useful. They include development of treatment guidelines, decisions within health care organizations, introduction of new medical technologies, reimbursement decisions and pricing decisions. In this review, DSU were considered potentially useful in making decisions about treatment choices and as weights in calculation of QALYs for CEA (economic evaluations). However, no study demonstrated how the DSU they elicited impacted on actual decisions.

This was not surprising considering Johannesson et al's (1996) observations that, "*even though the interest in incorporating both costs and health effects into public decisions concerning health and safety seems to be increasing, the impact of economic evaluations on actual decisions is still largely unclear.*" Despite the lack of evidence for developing countries, DSU are being used to make policy decisions in developing countries using DALYs (World Bank, 1993; Fox-Rushby, 2002). However, this may be an indication of the emerging currency in elicitation of DSU and the need to establish the validity, reliability and practicality of assessment of DSU before they can be put to actual decision making.

In the reviewed studies, there was not a clear demarcation between the two steps of measurement and valuation of disease states as less than half of them measured disease states or changes in disease states before eliciting the related DSU. In addition, most studies valued "own current health". While, by labeling, it may appear that subjects valued the same states, there was no way of telling if and how the state "own current heath" was linked to measurement instruments when used. Therefore, at present there lacks a transparency in defining the 'good' being valued, as links between what was being valued and measured are not explicit. In addition to making comparisons of health outcomes difficult for lack of descriptions of outcomes being compared, inability to separate measurement and valuation of health outcomes is a source of confusion in the literature. Therefore, to improve on measurement and valuation of disease specific utilities and to make them more relevant for economic evaluation and decision making, it would be useful to establish clear links between the two steps.

While the majority of studies provided no reasons for using a measurement instrument, reliability, validity and ease of use of instruments were provided more often as reasons for choice in those that did. The rest of the reasons were more specific to the instruments and conditions particularly with respect to disease specific measures. While this is indicative of the need for development of new instruments where none exists for a particular disease, it also points to the importance of assessing the measurement qualities of such an instrument. Considering the few studies providing reasons for choice of instrument, points to a state of

affairs where instruments are just being applied without justification, acknowledgement and or exploration of their suitability in different settings and situations.

Of the established valuation techniques, VAS, TTO and SG were the most commonly used, both singly and in combinations. No studies used PTO or ME. This reflects the views expressed in the wider valuation (Drummond et al. 1997; Gold et al. 1996) literature that these three methods are extensively used and the preferred choices for most researchers. While choice of the TTO and VAS was largely driven by simplicity and its extensive use, the choice of SG was largely driven by its theoretical grounding and reflection of decision making under risk and uncertainty. It is interesting to note that reasons governing the choice of instruments in practice were hardly ever based on the validity and reliability of the instruments, but rather on their popularity and ease of use. While these aspects are no doubt important in consideration of which instrument to use, only a few studies based their choice of instruments on how acceptable, practical, meaningful and understood, and valid and reliable the instruments were before applying them in different settings.

Most of the studies used patients as the raters, although many neither provided justification for choice of patients nor diseases considered. However, there was recognition that patients are the best source of information about their own health and quality of life because they have first hand experience and know the true implications of a particular health state. That SG was used more often with patients and health professionals than TTO or VAS, was perhaps an indication of recognition of risks involved in patient management and clinical decision making. That the TTO and VAS were more frequently used with general population than SG may be reflective of ease of use of these two techniques. The question of whose values to use in outcomes assessment for economic evaluation remains largely unresolved as there are merits and demerits of using various raters (Drummond et al. 1997; Johannesson, 1996; Gold et al. 1996; Ubel, et al. 2000; Williams, 1995; Brazier et al. 1999). However, Drummond et al (1997) and Gold et al (1996) suggest that the purpose and perspective of the task at hand should guide this choice.

Closely related to the issue of whose values is the question of which factors affect values. Several factors were assessed and can be categorized as socio-economic and demographic and factors related to illness and disease states. Relationships between age, gender and education were the most commonly assessed. Although the majority of studies found no significant relationships between factors considered and values, there were few reporting largely mixed findings. Hence, an unambiguous understanding of factors causing variation in values in different settings is still lacking. It was also not clear whether preferences for disease states varied as predicted by the various theories of choice reviewed in chapter 2. Also, investigation of issues like risk attitudes and whether axioms of theories of choice under certainty and uncertainty are confirmed or violated was not encountered and therefore constitutes information gaps.

Studies assessing factors affecting values failed to justify choice of factors, provide theoretical justifications and hypothesized relationships, as well as a priori expectation of the relationships between these variables and values. This constitutes a knowledge gap in understanding variation in values for disease states, especially because it was not clear how the relationships found related to theories underlying valuation of disease states.

Of the three major analytical tools used for assessing variation in values namely comparison of means, correlation and regression, none dominated and also a variety of test statistics were employed. While this might be a reflection considerations of the type of data being analyzed, it is a source of potential difficulties in comparing how different factors relate to values due to differences in methodologies. On the other hand, considering that the values are the units of analysis and are normally considered to be either on interval or ratio scale, it would be expected that analytical methods should be fairly similar.

Construct validity was the most commonly assessed form of validity for the valuation instruments. Empirical findings suggested that RS and TTO had construct validity for the contexts and diseases areas in which they were tested. There were mixed findings regarding construct validity of SG with some studies showing weak to very weak and others moderate to high correlations. The fact that these findings are based on only a few studies means that

they should be used cautiously but is also an indication of knowledge gaps in terms of understanding the validity of valuation instruments in obtaining disease specific utilities.

That construct validity was the most commonly investigated form of validity is probably a reflection of the fact that there is no single technique accepted as a gold standard against which criterion validity may be judged as well as practical difficulties in testing both criterion and content validity. It might also reflect the lack of explicit conceptual definitions of the concepts embodied in the valuation instruments against which content validity can be judged. While noting these findings, there is some concern about the way validity of valuation instruments is being judged in the literature. Correlations between valuation instruments and HRQL measurement instruments were used to judge validity of valuation instruments. This appears conceptually flawed because valuation instruments and measurement instruments are being treated as though they are measuring the same aspects and therefore can be compared with each other. As McLachlan et al (1999) notes, utility scales and HRQL measures are two approaches that measure different aspects of health outcome. They are compliments rather than substitutes and therefore it seems incorrect to validate one against the other. Utility scales are meant to measure preferences while HRQL instruments quantify and provide descriptive information of disease (health) states. These findings reinforce the sparseness of knowledge about the various forms of validity of valuation instruments when used to assess DSU. Therefore, there is need to establish the content and construct validity of valuation instruments further especially in new settings.

There is a paucity of studies assessing reliability and practicality of valuation instruments when used to assess DSU. In terms of practicality, there was also a lot of variability in the indicators used to assess practicality and no instrument was subject to these indicators uniformly. This makes it difficult to compare them adequately and or judge their practical performance. As such, more research work comparing these valuation techniques with similar criteria is required to contribute to an understanding of their performance in new settings.

Although there is dearth of studies assessing validity, reliability and practicality of valuation instruments for disease specific utilities, these aspects have been found to be satisfactory for RS, TTO and SG in North America and Europe amongst generic health states. However, performance on validity, reliability and practicality varies widely even in these settings and therefore cannot be used as a case for no further testing. Furthermore, disease specific measures concentrate "on the symptoms, complaints and disruptions in life that are specific to the disease, are tailored to specific needs, and are more sensitive than generic instruments" (Willliams and Wood-Dauphinee in Fox-Rushby, 1994). Therefore, we cannot base content validity of disease specific utilities on generic utilities because the domains of content are different. Construct validation is an on-going process that must be conceived within a theoretical context so as to assess how the measure performs in accordance with theoretical expectations (Carmines and Zeller, 1979). As Carmines and Zeller (1979) note "construct validity is not established by confirming a single prediction on different occasions or confirming many predictions in a single study. It requires a pattern of consistent findings involving different researchers, using different theoretical structures across a number of different studies." Streiner and Norman (1995) argue that "every time a scale is used in a new context, or with a different group of people, it is necessary to re-establish its psychometric properties". To the extent that these instruments will be used in a new context, new disease and new population raises cross-cultural issues and justifies further testing of their psychometric properties.

Cross-cultural issues in application of both measurement and valuation instruments has received little attention especially in low-income countries like Kenya. Of particular interest in the Kenyan study was that the instrument was not pre-tested neither its performance assessed in its traditional form before adapting it. It is thus not explicitly clear how the adaptation process was carried out and why. Therefore, performance and applicability of these instruments in these settings remains largely unknown.

Content and criterion validity of VAS, TTO and SG were subject to virtually no testing in disease specific utilities. Although the studies however tested construct validity, lack of content validity may cast serious doubts on claims of construct validity. Herdman et al

(1998) have suggested testing conceptual equivalence as a first step towards facilitating justification of use of instruments in different settings and laying the framework for claims on any other form of equivalence, including construct validity which falls under measurement equivalence. These concerns notwithstanding, testing of construct validity was largely in developed countries context, implying paucity of knowledge on both content and construct validity of the instruments in low income settings. This constitutes a methodological knowledge gap in use of the existing valuation instruments in a low-income country.

CHAPTER 5

DEVELOPMENT OF A MEASURE FOR SCHISTOSOMIASIS MANSONI DISEASE STATES IN KENYA

5.1 INTRODUCTION

This chapter presents the methods used in the development of a tool to measure the impact of schistosomiasis disease states on HRQL in Kenya. In chapter 3, it was noted that none of the existing generic index-based HRQL measurement instruments had been applied in Kenya. Chapter 4 also established that no valuation for schistosomiasis disease states existed. The challenges of developing and or adapting the existing instruments were also noted in chapter 3, and this resulted in the decision to develop a disease specific questionnaire that could capture the HRQL impacts of Schistosomiasis Mansoni in Kenya. This chapter thus describes the development of the questionnaire for use in a Kenyan community.

The chapter is organised as follows. The next sub-section section presents approaches to questionnaire development. Section 5.2 presents a review of Schistosomiasis Mansoni literature, highlighting how Schistosomiasis Mansoni may affect HRQL. Sections 5.3 and 5.4 present methods for the development and description of the tool for assessing the HRQL impacts of Schistosomiasis Mansoni in Kenya.

5.1.1 Approaches to Questionnaire Development

The essence of a measurement tool is that it taps all the relevant aspects of the phenomena being measured. It is therefore vital that the tool contains items that tap different aspects of the phenomena, in this case the experience of symptoms as well as their impact on HRQL. There are several avenues through which items for a questionnaire can be generated and choice of which strategy to adopt may be influenced by resource and time availability (Guyatt, 1995).

The standard and widely accepted practice in questionnaire development has been the psychometric approach (Cano, 2001). The approach proceeds through various steps each feeding into the next. The initial steps include literature reviews and reviews of existing instruments (if any), expert opinions and consultations with patients and health professionals as well as in-depth qualitative interviews with patients and experts. This process helps in item generation and development of a conceptual model detailing the various domains to be included in such an instrument. A second step involves pre-testing the long form questionnaire on a small sample (15-25) of the target population with a view to identify items that respondents have difficulties understanding, or interpret differently than is intended. It also helps to clarify ambiguities in the wording of items, confirm appropriateness, and determine acceptability and completion time. This process leads to modification of the questionnaire through the third step that is item reduction. The criteria for item reduction normally includes testing for missing data, item redundancy, ceiling and floor effects and item-total correlations (Cano, 2001). Item reduction leads to development of scales by method of factor analysis and consequently the final step of testing the instrument for validity, reliability, responsiveness, sensitivity and acceptability (Cano, 2001).

Fayers and Hand (2002) distinguish between the traditional psychometric and clinimetric approaches to questionnaire development, the differences being based on whether the questionnaire contains causal[36] and or indicator[37] variables. Providing several examples where the psychometric approach led to dropping of most plausible and important items, they argue that where an instrument contains causal variables the psychometric approach is unsuitable and the clinimetric approach more suitable. "The clinimetric approaches are based on a deliberate choice of what variables to include and, in the absence of an underlying model, a deliberate choice of how these variables should be combined" (Fayers

[36] Causal variables are variables that are part of the definition of the concept being measured. Hence if the variables are present, then the concept in question is present. Use of causal variable does not assume that the concept exists.

[37] Indicator variables are merely aspects of the concept being measured and they do not alter or influence the concept. They concept is assumed to exist and the variables just indicate whether they have a relationship with the concept being measured.

and Hand, 2002: p.241). Subjecting clinimetrically developed questionnaires to psychometric testing is unsuitable as they contain causal variables (Fayers and Hand, 2002). Fayers and Hand (2002) observe that disease specific QOL instruments can be expected to contain a higher proportion of causal items and hence a clinimetric approach should be followed in their development.

The process of developing and testing the psychometric properties of a questionnaire is long, intensive and time and resource consuming, not to mention the requirements of active involvement of experts from different disciplines. In fact it can constitute a whole Ph.D study (see Cano, 2001). While the aim of this thesis was not questionnaire development, it was deemed inappropriate to use the existing generic health measurement instruments on account of cross-cultural issues and that their psychometric properties had not been determined in the Kenyan setting. Also of concern from the outset was whether the existing generic instruments would be sensitive and responsive enough to capture the concerns of people suffering from Schistosomiasis Mansoni. Hence, it become necessary to search for a reliable and valid disease specific measure in Schistosomiasis Mansoni but none was found, culminating in the decision to develop a Schistosomiasis Mansoni disease specific measure that would then be used to measure the impact of schistosomiasis in Kenya. Due to time and resource limitations, and concerns for the size of the thesis, the traditional methods of instrument development were not strictly followed and this may be considered as a drawback of this study.

5.2 SCHISTOSOMIASIS MANSONI AND ITS IMPACT ON HEALTH STATUS

This literature review aims to reveal important areas where Schistosomiasis Mansoni may affect HRQL. This information was used in the development of a tool to assess the impact of Schistosomiasis Mansoni on patients and community members in Mwea. The review examines the transmission mechanism, epidemiology, control strategies, impact of schistosomiasis on HRQL, and outcome measures used to date in economic evaluations of Schistosomiasis Mansoni interventions.

5.2.1 Search Strategy

Medline (searched from 1966–1999) and *Health Star* (searched from 1966–1999) databases were searched to identify literature relevant to the topic of interest. The Cochran library and WHO web page was also searched for reviews on Schistosomiasis Mansoni. Search terms for literature on Schistosomiasis infestations included intestinal worm*; intestinal helminth*; *schistosoma mansoni; bilharzia*;* control strateg*; intervention strateg*; effective*; cost*; cost effective*; outcome*; economic evaluation; parasitic infection* and Kenya*. The terms were combined as appropriate using the boolean operators such as 'and' and 'or'. The search was limited by language of publication (English). For every search, the abstracts were examined to select relevant articles. Some references mentioned in the papers identified were followed up manually as well as those suggested by experts in the areas. Relevant chapters in published textbooks and theses were also reviewed.

5.2.2 Public Health Significance, Transmission and Epidemiology

Public Health Significance

Schistosomiasis is the second most prevalent tropical disease after malaria (WHO, 1998) and is a leading cause of severe morbidity (WHO/CTD, 1999). Globally, 600 million people are at risk, with more than 200 million infected, 120 million[38] symptomatic and 20 million suffering severe consequences from the disease. In terms of mortality, more than 250,000 deaths annually are estimated to be associated with schistosomiasis (Awasthi et al. 2003; WHO/CTD, 1999; WHO, 1998). It is estimated that 85% of all the people infected with schistosomiasis are in Africa (Awasthi et al. 2003). In terms of the socio-economic and public health importance in the tropical and sub-tropical areas, it is second only to malaria (Stephenson and Holland, 1987). Of the three common forms of schistosomiasis, S. Haematobium infects at least 78 million persons, S. Mansoni an estimated 57 million persons and S. Japonicum 69 million persons (Stephenson and Holland, 1987).

[38] WHO, 1998 refers to this figure as those with severe activity limitation which is permanent and long term.

Disease due to schistosomiasis depends on the species (*S. Haematobium (*SH*), S. Mansoni (*SM*), S. Japonicum (*SP*), S. Intercalatum (*SI*) and S. Mekongi* SM*))* a person is infected with and the intensity of past and current infection (WHO/CTD, 1999). S. Japonicum, and S. Mekongi are prevalent in 7 Asian countries and the Pacific region. S. Intercalatum is found in 10 African countries while S. Heamatobium is found in 54 countries in Africa and Eastern Mediterranean. S. Mansoni is found in 52 African countries, the Caribbean the Eastern Mediterranean and South America (WHO/CTD, 1999). The WHO estimated that helminthic infections represented more than 40% of the disease burden from all tropical disease excluding malaria (Awasthi et al. 2003). S. Mansoni is one of the most prevalent and widespread of the three species (WHO/CTD, 1999).

Transmission

The life cycle of Schistosoma Mansoni has two hosts, a freshwater snail (Biomphalaria genus) and a human. Lack of good sanitary and hygienic practice leads to contamination of the environment, which aggravates the transmission of the parasite (Stephenson and Holland, 1987). The eggs of schistosomes in the excreta of an infected person open on contact with water in the presence of sufficient sunlight and warmth and release a parasite, *miracidium.* To survive, this parasite must find a fresh water snail intermediate host which they must enter within 24 hours or die (WHO/CTD, 1999: Stephenson and Holland, 1987). Once in the host (fresh water snail), the miracidium becomes a first stage sporocyst, which divides producing thousands of new parasites (Cercariae), which are then excreted by the snail into the surrounding water. The multiplication within the snail takes about 4 to 8 weeks (Stephenson and Holland, 1987). Cercariae released by the snail penetrate an individual's skin within seconds and assisted by enzymes, the parasite migrates through the epidermis into the blood stream where they continue their biological cycle (Saconato and Atallah, 1999).

After penetrating the skin they become schistosomula and are carried passively in the vascular system from the subcutenous tissue to the right heart and then on to the lungs where remain and grow for about three days. They then proceed to the hepatic portal system of the liver through an uncertain route (Stephenson and Holland, 1987). Once in the

liver they become sexually mature and mate and then migrate to the vesical or mesentric veins and begin egg production. It takes approximately 30-45 days for the parasite to be transformed into a long adult worm (WHO/CTD, 1999). The adult worms are 1-2 centimetres in length and occur in pairs, male-female. Adult S. Mansoni live in the portal system, primarily in the superior mesenteric veins as well as the blood vessels lining the intestine. The female worms produce 200-2000 eggs per day over an average of 5 years (WHO/CTD, 1999). Through a combination of enzymatic secretions and peristalsis the eggs pass out of the blood vessels through the tissues and into the lumen of the gut to leave the body in faeces (Stephenson and Holland, 1987). Only a half of the eggs are excreted in the faeces (to continue the transmission cycle), with the rest staying in the body damaging other vital organs. These eggs are trapped in the tissues and provoke pathology, particularly the formation of granulomas (Stephenson and Holland, 1987). It is the eggs and not the worm which cause damage to the intestine, the bladder, spleen and liver (WHO/CTD, 1999). The mean life span of an adult worm is between 3 and 8 years (Stephenson and Holland, 1987: Chan et al. 1995: WHO/CTD, 1999).

Epidemiology

The distribution of schistosomiasis varies widely (WHO, 1993) within and across countries, endemic areas in a country as well as within individuals at risk of the infection. Epidemiological studies (Guyatt, 1998: Chandiwana and Christensen, 1988: Chandiwana et al. 1988; Butterworth et al. 1991, 1994) have established that prevalence and intensity of infection with schistosomiasis as with other intestinal helminths is highly age-dependent, with some age groups harbouring the most intense infections. It has also been established (Warren et al. 1993: Anderson and Medley, 1985: Guyatt et al. 1993: Muthami et al. 1995: Chandiwana et al. 1988: Gryseels et al. 1989: Butterworth et al. 1994 and 1991) that within an infected community there are a few wormy people and the majority have few worms, which is essential for identifying the populations at risk of morbidity.

Much of the burden from schistosomiasis falls on children, particularly those of school age (5-19) who have the highest prevalence and intensity of infection (Stephenson, 1993: WHO, 1993: Guyatt and Evans, 1992: Chan et al. 1994. Evans and Guyatt, 1997). Children

constitute a high proportion of population in most developing countries, implying that they bear a substantial burden of the disease as well as contribute most to its transmission. As Hatz et al. (1990) states, groups with the highest prevalence and intensity and consequently at the highest risk of developing pathological lesions are children and adolescents from age 6-20 years of age. Studies in Kenyan schistosomiasis endemic areas report prevalence among school children (Jaoko et al. 1996: Butterworth et al. 1991, Kloos et al. 1987: Hunter et al. 1993) in the range of 70% to 90% and among adults (Muthami et al. 1995: Kirigia, 1994) 31% to 64%. Prevalence of hepatomegaly reported for Kenya range between 10% to 40% (Butterworth et al. 1991 and 1994). This implies that both children and adults are at risk in these communities.

5.2.3 Control Strategies

The basic strategies used in treatment, prevention and control of schistosomiasis include chemotherapy, water supply and sanitation, health education and snail control (molluscicide). These four components constitute what WHO (1993) regards as a feasible and effective strategy for morbidity control. However, the use of these strategies at the country level depends on the epidemiological situation. Chemotherapy plays a central role in any strategy of control and dramatically reduces morbidity in the short term (WHO, 1993). Water supply and sanitation and health education programs are very expensive to implement but they yield long term gains (Warren et al. 1993).

5.2.4 Impact on Health Status

Although S. Mansoni infection is largely asymptomatic *"....this insidious and chronic disease lacks the drama usually associated with other spectacular infections.... The parasite relies on co-existence rather than elimination of its host, and this results in chronic debility.....with underestimated overt morbidity and mortality. But the damage done to the individual and the community is much more than meets the eye. Our present*

difficulties relate mostly to the measurement of the invisible damage done", Farooq (1964) in (Stephenson and Holland, 1987).

Morbidity due to schistosomiasis is not caused by worms per se, but by the accumulation of eggs in the capillary beds, consequent formation of granuloma and the simultaneous immune response to the eggs. This results in reversible and irreversible damage of organs such as the intestine, liver, spleen, bladder and kidneys (Stephenson, 1993: Medley and Bundy, 1996). Various studies (Guyatt, 1998: Medley and Bundy, 1996: WHO, 1985) have attempted to classify morbidity into acute, early and late chronic disease. Intensity of infection is a major determinant of disease (Saconato and Atallah, 1999) and hence the mean egg count is a useful indicator of morbidity (WHO, 1993). However, the relation of morbidity to disease may be to the past, rather than current levels of worm burden. Schistosomal disease should be viewed as a progression through different stages (asymptomatic, mild, moderate, severe, very severe, comatose and dead) that become progressively difficult to resolve through treatment (Kirigia, 1998)[39]. As one progresses through these severity stages, it is suspected that one's ability to function becomes progressively diminished.

In addition to causing acute and chronic morbidity and mortality, schistosomiasis impacts on nutritional status, growth and physical fitness, educational attainment and cognitive abilities as well as work productivity[40].

Acute morbidity: Acute morbidity is associated with intestinal signs and symptoms-colicky abdominal pain, bloody diarrhoea and blood in stools (Gryseels, 1992). These symptoms may have a chronic or intermittent character, and the intensity ranges from mild discomfort to severe or sometimes fatal dysenteric syndromes (Gryseels, 1992). In many cases, there is a significant association between the presence of infection and of diarrhoea and particularly bloody diarrhoea and or presence of blood in stools (Gryseels, 1992). Severe chronic or intermittent dysenteric syndromes clearly put a heavy burden on the

[39] See Kirigia (1998) for seven main severity states and their accompanying clinical symptoms.
[40] See appendix 5.1.

affected communities (Gryseels, 1992). Patients with intestinal symptoms may remain otherwise asymptomatic until the disease is well advanced or haematemesis occurs (Abdel-Wahab et al. 1993). The acute form of schistosomiasis morbidity is assumed to be proportional to current infection intensity and resolves after treatment approximately in three months (Medley and Bundy, 1996). Most acute morbidity is associated with heavy infection[41].

Chronic Morbidity: Severe disease due to schistosomal infection follows after many years (often after 20 years of age) of silent or mildly symptomatic infection (WHO, 1993). In intense infections and with continuing exposure, more important pathology develops partly as a result of build-up of fibrous tissue remaining after granuloma resolution (Medley and Bundy, 1996). In the chronic severe form of this disease, scarring and fibrosis result in enlargement of the liver and the spleen. The ensuing portal hypertension is the cause of death from massive haematemesis from oesophageal varices. Hepatic fibrosis and portal hypertension are life threatening and irreversible in advanced disease (WHO, 1993: Medley and Bundy, 1996: Gryseels, 1992). Patients with severe acute schistosomiasis may require admission to hospital for diagnosis and treatment (WHO, 1993). Mortality from schistosomiasis has been poorly documented in most endemic countries (WHO, 1993) which means that mortality due to schistosomiasis continues to be underestimated.

HRQL impacts: Schistosomiasis has harmful effects on nutritional status and growth rate, educational attainment in terms of intellectual capacity and school attendance, labour or work productivity, physical fitness and activity (see appendix 5.1). The harmful effects of schistosomiasis on the growth, development and health status of school-age children (5-19 years) are now recognised to be greater than was previously suspected (WHO, 1993: Stephenson, 1987, 1989). Poor growth in children and weight loss in adults can result from a reduction in appetite (Stephenson, 1993). Educational impairment reduces investment in human capital of children with long term implications on personal and societal well being (Kirigia, 1994). Watkins (1996) found that a higher worm burden was associated with poorer performance measured in terms of tests of reading and vocabulary, but cautioned

[41] See appendix 5.2 for various classifications of light, moderate and heavy infections.

that effects of worms may be modest compared to social and economic factors. Fatigue, reduced physical fitness, weakness, lassitude, muscle and abdominal pain, nausea and vomiting, decreased food intake (WHO/CTD, 1999: WHO, 1993: Stephenson, 1993: Stephenson and Holland, 1987: Warren et al. 1993) may lead to a reduction in an individual's capacity to carry on their activities of daily living such as work and leisure, which in a vicious cycle leads to less productivity, poor health and quality of life.

Although the causation mechanisms have not been clearly established, it appears that most of these harmful effects are driven by the various signs and symptoms such as blood loss which causes anaemia and subsequently impairs nutritional status, diarrhoea, vomiting, nausea, anorexia, decreased food intake which leads to nutrient loss and impaired growth and hence inability to take advantage of educational and other opportunities in life. Evidence of the impact of Schistosomiasis Mansoni on the physical fitness, productivity, anthropometry and school performance is rather conflicting with some showing no impact while others show impact (Gryseels, (1992: Tanner, 1989: Stephenson, 1993), thereby making it difficult to substantiate claims of impacts (WHO, 1993), though highly suspect.

5.2.5 Outcome Measures used in CEA of Schistosomiasis Mansoni Interventions

Reflecting on the impact of schistosomiasis on the health of the victims discussed above, the insidious and chronic nature of the infection and disease development coupled with difficulties related to measuring disability and handicap caused by the disease, assessment of health related quality of life is suggested as an alternative to assess the burden of disease due to schistosomiasis. Commonly used effectiveness or outcome measures such as prevalence and intensity of infection, number of people cured or case years prevented say nothing about the quality of life and the desirability of such life as lived by those infected. These measures also fail to reflect the impact of schistosomiasis on nutritional status, growth and development, physical fitness and activity, educational attainment and cognitive abilities as well as on work productivity, which subsequently affect health and quality of life. We suggest that a HRQL approach to assessing such affects may overcome some of

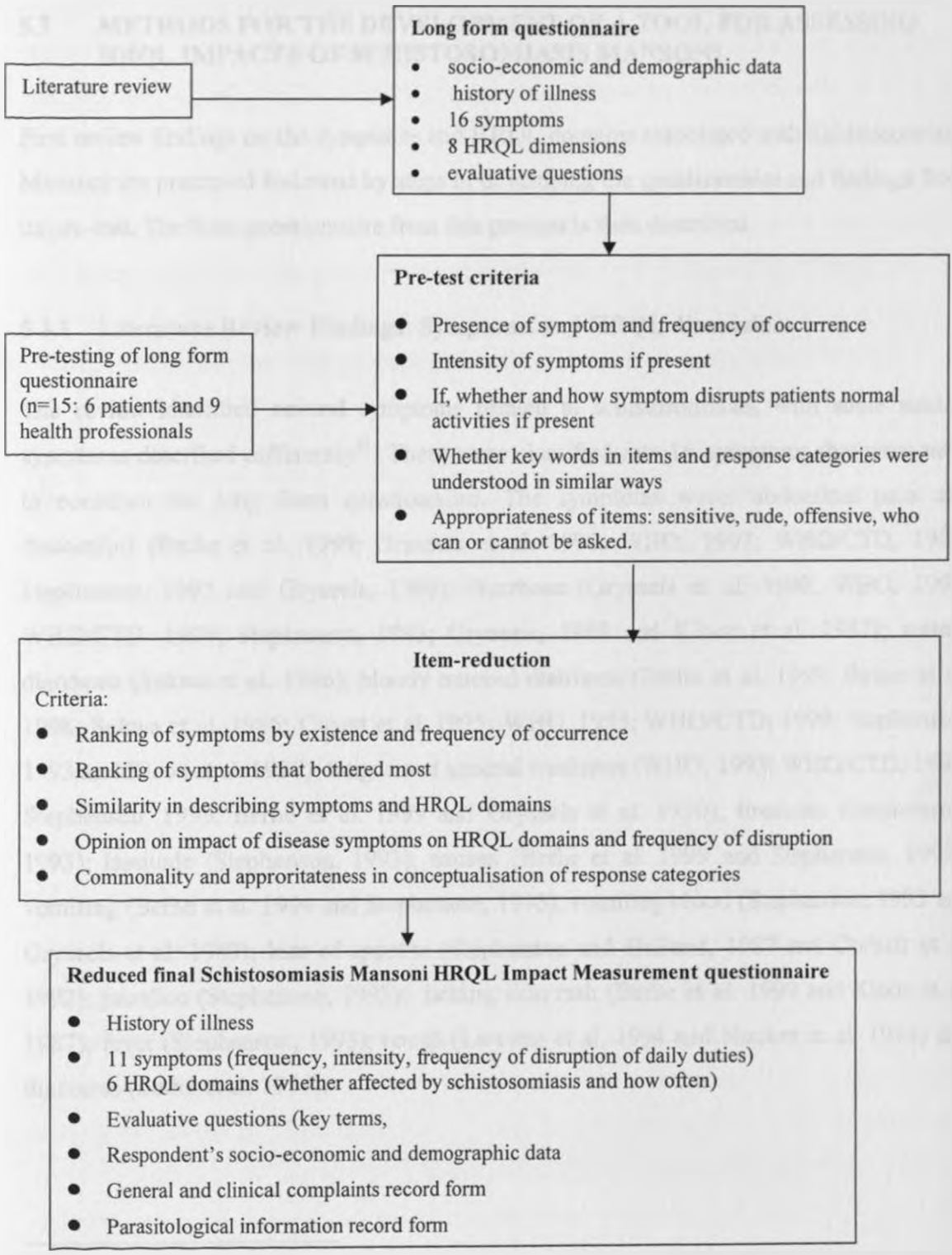
the difficulties experienced assessing the significance and impact of schistosomiasis on affected communities.

5.2.6 Summary

Although evidence is inconclusive, it is generally felt that schistosomiasis infection has adverse effects on health status and subsequent quality of life of those infected. This is through its effects on nutrition, growth and development, physical fitness, educational attainment and cognitive abilities as well as on work productivity. This is in addition to the effect of experiencing the symptoms. As such schistosomiasis is a social problem whose impact and significance on health and quality of life is neither fully appreciated nor adequately measured to reflect the concerns of the sufferers.

Most research in schistosomiasis in Kenya has concentrated on epidemiological studies and recently on control strategies (Katsivo et al. 1993) such as chemotherapy. In spite of these advances, only Kirigia (1994) has attempted to assess the quality of life the people with schistosomiasis experience in Kenya. Measurement of the public health impact of the disease has rarely if ever, gone beyond prevalence and intensity of the various indicators of morbidity, to explore whether and how infection with Schistosomiasis Mansoni impacts on people's quality of life and how people perceive and value such an impact. Due to the insidious and chronic nature of schistosomiasis and the difficulties associated with attributing death and disability to the infection, I propose that health related quality of life is an approach worth developing for assessing and expressing the HRQL impacts associated with schistosomiasis infection in measuring outcomes of intervention programs. In doing so, the views and perceptions of the community on health and quality of life should be incorporated in to the measure and the measurement properties of the instrument established. The next section deals with the development of a tool for assessing HRQL impacts of Schistosomiasis Mansoni following the steps presented in figure 5.1.

Figure 5.1: Flow chart of steps in developing questionnaire



5.3 METHODS FOR THE DEVELOPMENT OF A TOOL FOR ASSESSING HRQL IMPACTS OF SCHISTOSOMIASIS MANSONI

First review findings on the symptoms and HRQL domains associated with Schistosomiasis Mansoni are presented followed by steps in developing the questionnaire and findings from its pre-test. The final questionnaire from this process is then described.

5.3.1 Literature Review Findings: Symptoms and HRQL dimensions

The review identified several symptoms related to schistosomiasis, with some similar symptoms described differently[42]. These were classified into 16 symptoms that were used to construct the long form questionnaire. The symptoms were: abdominal pain and discomfort (Berhe et al. 1999; Gryseels et al. 1990; WHO, 1993; WHO/CTD, 1999; Stephenson, 1993 and Gryseels, 1989); diarrhoea (Gryseels et al. 1990; WHO, 1993; WHO/CTD, 1999; Stephenson, 1993; Gryseels, 1989 and Kloos et al. 1987); watery diarrhoea (Sukwa et al. 1986); bloody mucoid diarrhoea (Berhe et al. 1999; Boiser et al. 1998; Sukwa et al. 1986; Guyatt et al. 1995; WHO, 1993; WHO/CTD, 1999; Stephenson, 1993 and Kloos et al. 1987); fatigue and general weakness (WHO, 1993; WHO/CTD, 1999; Stephenson, 1993; Berhe et al. 1999 and Gryseels et al. 1990); tiredness (Stephenson, 1993); lassitude (Stephenson, 1993); nausea (Berhe et al. 1999 and Stephenson, 1993); vomiting (Berhe et al. 1999 and Stephenson, 1993); vomiting blood (Stephenson, 1993 and Gryseels et al. 1989); loss of appetite (Stephenson and Holland, 1987 and Corbett et al. 1992); jaundice (Stephenson, 1993); itching skin rash (Berhe et al. 1999 and Kloos et al. 1987); fever (Stephenson, 1993); cough (Lawless et al. 1994 and Nookes et al. 1994) and dizziness (Berhe et al. 1999).

[42] For example, to refer to abdominal pain and discomfort, terms such as abdominal cramps, abdominal discomfort, colicky abdominal pains, stomach disorders, were used. Similarly, diarrhoea was described using terms such as watery diarrhoea, bloody mucoid diarrhoea and dysentery while tiredness was described as lassitude, fatigue and general weakness.

For purposes of clinical assessment for the patients' survey, clinical signs were also identified. These included hepatomegaly (Sukwa et al. 1986; Gryseels et al. 1990; WHO, 1993; WHO/CTD, 1999; Stephenson, 1993; Gryseels et al. 1989 and Kloos et al. 1987); splenomegaly (WHO, 1993; WHO/CTD, 1999; Stephenson, 1993; Gryseels et al. 1990; Gryseels et al. 1989; Dominguesnet al. 1990 and Kloos et al. 1987); anaemia (WHO, 1993; WHO/CTD, 1999 and Stephenson, 1993); oedema (Stephenson, 1993), ascites (Stephenson, 1993), hepatic coma (Stephenson, 1993), pulmonary hypertension (Stephenson, 1993) and periportal fibrosis (Stephenson, 1993). Since clinical signs can only be detected through clinical diagnosis by clinician, these were not subjected to pre-testing, but were included in the final questionnaire section on clinical assessment.

Several studies have documented the harmful effects of Schistosomiasis Mansoni and other intestinal worms on peoples' lives (see appendix 5.1). Although literature does not establish clearly the causation mechanisms through which schistosomiasis infection may affect health related quality of life, there are indications that it does. The argument in this thesis is that the most comprehensive way that impact of schistosomiasis on peoples' lives can be assessed is through documenting its impact on health related quality of life of those infected and subsequently valuing those impacts. This arises from the fact that previously used measures say nothing about the quality of life lived with the infection and desirability of such a life.

To identify HRQL dimensions that schistosomiasis may have an impact on, literature was gleaned for effects or indications that could be associated with health related quality of life domains consistent with WHO's (1993) concept of health. These were classified into physical, social and mental dimensions. In terms of the physical dimension, indicators included growth, physical fitness and work capacity (Watkins et al. 1996; Connolly et al. 1993; Stoltzfus et al. 1997a; Stoltzfus et al. 1997b; Stephenson et al. 1994; Stephenson et al. 1990; Adams et al. 1994; WHO, 1993; WHO/CTD, 1999 and Stephenson, 1993) and physical appearance such as swollen abdomen (Kloos et al. 1987). Kirigia (1994) included mobility, livelihood activities and self care in health states he valued in Mwea. Although Kirigia (1994) included the domain of social participation, social and mental dimensions

were not addressed in any of the reviewed studies. However these were included in the first draft questionnaire to be pre-tested in order to ascertain whether or not they are concerns to people suffering from schistosomiasis in Mwea. The sample on which the questionnaire was pre-tested is described next.

5.3.2 The Sample

A convenient sample of nine health professionals with practice and experience of Schistosomiasis Mansoni experience was selected. This comprised of three doctors one of whom was based in Mwea and two in Nairobi (DVBD), two clinical officers and four nurses all based in Mwea. The address and information of patients recently treated of Schistosomiasis Mansoni at two GOK and one private Medical Centre based in Mwea was obtained and used to trace six patients in their homes, where consent and permission to interview was obtained. Hence, the pre-test sample comprised of 15 conveniently selected respondents.

5.3.3 Steps followed in developing the questionnaire

The study followed a clinimetric approach (Fayers and Hand, 2002) to generate and choose items that ended up in the Schistosomiasis Mansoni HRQL questionnaire. The concern was to include symptoms and assess whether and how they contribute to activity disruption and consequently whether and how this impacts on HRQL. A long-form questionnaire was constructed and subsequently pre-tested on schistosomiasis experts and patients at health facilities in Mwea. Following the pre-test the questionnaire was further adjusted and refined based on the views of the patients and experts and following a set criteria (see figure 5.1). This led to the construction of final Schistosomiasis Mansoni HRQL questionnaire that was used in the empirical work on measurement of schistosomiasis disease states in Mwea.

The long form questionnaire assessed whether the symptom was present, its frequency and intensity, whether it disrupted normal activities, in what ways and how often. Patients were asked to state if they had experienced the symptoms while experts were asked their

opinions regarding patients' complaints. Also assessed were opinions on the most bothersome symptoms and those best representing illness with bilharzia. With regard to HRQL domains, information was collected on whether the domains were affected due illness with bilharzia, and if so, the extent and ways in which it was affected. Finally there were evaluative questions regarding meanings attached to response categories and key words as well as appropriateness in terms of who can and cannot be asked the questions.

The long form questionnaire that was constructed for pre-test included:

- Background information (age, gender, marital status, and educational attainment, profession, occupation and income level).
- Illness and health problems experienced currently and in the last four weeks.
- Questions on frequency, intensity and activity disruptions for each of the sixteen symptoms outlined above.
- Open-ended questions to assess which symptoms patients thought bothered them most and description of how life and its quality is when one has schistosomiasis (bilharzia).
- In choosing the HRQL domains, the study adopted WHO's (1993) concept of health and then considered how schistosomiasis literature fitted. It did for the physical and social domains, with no studies addressing the mental domain. Due to the variation in the way physical domain indicators were presented in the literature, I decided to use the following domains consistent with different references in the literature; mobility (Stoltzfus, 1997a; Stephenson 1990; Stephenson 1994; Adams, 1994), activities of daily living (Stephenson 1990; Stephenson 1994; Adams, 1994; Parker, 1992), work productivity (Wakins, 1996; Stephenson 1990; Stephenson 1994; Adams, 1994) and feeling of strength and energy in body (Lawless, 1994; Hadju et al, 1996; Stoltzfus, 1997a; Stephenson 1994; Adams, 1994) to represent the physical domain. Participation in social functions (Kirigia, 1994; Adams, 1994; Parker, 1992) represented the social dimension. Owing to lack of studies addressing the mental domain and to reflect the WHO's (1993) conception of health, I included feelings of tension, hopelessness as well as despair, worry and anxiety and thoughts to represent the mental dimension. These eight health related quality of life domains

were assessed in terms of whether respondents thought schistosomiasis affected them, how they thought they were affected and how often they were experienced (from none of the time to most of the time).

- A section on evaluation questions was included to assess the meanings attached to some key words and phrases, their appropriateness in terms of who can and cannot be asked certain types of questions. Key words assessed included phrases like 'very rarely', 'rarely', 'often', 'very often', 'mild', 'moderate', 'severe' 'very severe', 'a little, some and most of the time' that were used as response categories to assess frequency and intensity of symptoms and extent of impact on HRQL dimensions. Respondents were also asked to describe what they referred to in key terms such as 'illness', 'health problems' 'daily activities', 'activities of daily living', 'normal activities', 'disruption', 'bother', and 'health state'. Respondents were also asked to point out any questions they thought could not be asked to a young boy or girl, young man or woman and old man or woman. In addition they were asked to say if they had found any questions sensitive, rude or offensive.

The criteria used in pre-testing the questionnaire was whether a symptom was reported as present as well as the frequency and intensity of it occurrence. This was used as an indicator of how important the symptom was. If the symptom was present, another criterion was whether and how it disrupted normal activities. Opinions on whether key words in items and response categories were understood in similar ways were also used as pre-test criteria in addition to appropriateness of items in terms of being sensitive, rude, offensive and who can and cannot be asked.

To facilitate item reduction, the following criteria were used. Ranking of symptoms by existence and frequency of occurrence and symptoms that bothered most. These two criteria were used to drop symptoms that ranked low in being present and low in importance as indicated by bother in Schistosomiasis Mansoni patients. Symptoms and HRQL dimensions were also dropped if they were described in similar local terms as this implied redundancy. Opinions on impact of symptoms on HRQL were also used as these indicated if, whether and how the various HRQL dimensions were affected and by which

symptoms. Finally, in regard to response categories, the criteria of appropriateness and commonality of conceptualization of response categories was used. Hence if a response category was inappropriate in describing the response, an appropriate description was adopted following pre-test suggestions. Similarly, major differences in conceptualizing what a response category meant to different respondents implied that the response categories were not measuring similar concepts and hence an appropriate response that which would have a similar meaning to most if not all respondents was used.

5.3.4 Questionnaire Pre-test Findings

5.3.4.1 Symptoms

Reports of experience of the symptoms in the last two weeks by patients and of how often patients complained of the symptoms by experts produced rankings that were both similar and different. The ranking of symptoms obtained from patients and experts was slightly different with two patients reporting having watery, bloody diarrhoea and none having diarrhoea while all experts reported that patients complained of various forms of diarrhoea. Symptoms like jaundice, vomiting and vomiting blood, itching, fever and cough were reported as less likely to be found in patients by three to six experts. Similarly, only two patients reported having cough and jaundice while one reported itching and vomiting and none had diarrhoea and vomiting blood. Except for fever, which was reported by all patients, there was thus some agreement in opinion between patients and experts regarding the unlikely presence of jaundice, vomiting, vomiting blood, itching and cough.

Both patients and experts were asked to state the symptoms that were most bothersome. Again there was slight difference in opinion. Experts regarded bloody mucoid diarrhoea, abdominal pain and discomfort and tiredness as the most bothersome symptoms. On the other hand, patients reported that abdominal pain and discomfort, dizziness and loss of appetite as the most bothersome. Other symptoms were also mentioned but less often. When asked which symptoms best represent bilharzia, one to two patients mentioned

bloody diarrhoea, abdominal pain and discomfort and nausea while all the experts mentioned bloody diarrhoea. However, one expert noted that although bloody diarrhoea is the best indicator, microscopic examination is the best determinant because there are other diseases and conditions such as shigella and dysentery that cause bloody diarrhoea.

From the above findings, there was agreement of opinion that abdominal pain and discomfort, bloody diarrhoea, tiredness, dizziness, loss of appetite and nausea were considered more likely to occur and also most bothersome. There was a slight difference of opinion regarding occurrence of diarrhoea and watery diarrhoea, vomiting and itching skin rash between patients and experts, except for fever. This difference of opinion could be attributed to either patients' reluctance to talk about their diarrhoea (embarrassing) or perhaps a misguided but widely held view amongst experts that bloody mucoid diarrhoea is commonly found in Schistosomiasis Mansoni patients. Whichever the case, this is an area requiring further exploration to clearly unravel patients and experts views regarding symptoms commonly associated with the disease. During the pre-test, both patients and experts considered fatigue and general weakness, lassitude and tiredness to mean the same thing. Tiredness was chosen to represent them. Hence, the eleven symptoms chosen to be included in the final questionnaire were abdominal pain and discomfort, bloody diarrhoea, tiredness, dizziness, loss of appetite, nausea, diarrhoea, watery diarrhoea, vomiting, itching skin rash and fever.

5.3.4.2 HRQL Dimensions

There was considerable agreement between patients and experts in opinions regarding whether bilharzia could affect the eight HRQL dimensions. Except for feelings of tension, hopelessness and despair, the majority of experts (7-9 out of 9) agreed that mobility, activities of daily living, work productivity, feelings of strength and energy in body, social participation, feelings of anxiety and worry and thoughts about infection were affected by bilharzia. The majority (5-6 of 6) of patients felt similarly that Schistosomiasis Mansoni affected most of the dimensions. However, both experts and patients felt that feelings of tension, hopelessness and despair and thoughts about infection were very similar to feelings of anxiety and worry and therefore the term worry and anxiety was used to represent them

to avoid redundancy of items. Hence, the final questionnaire contains six HRQL dimensions.

5.3.4.3 Patients' and experts' opinions on impact of Schistosomiasis Mansoni on HRQL

Opinions of Symptoms

Symptoms reported as causing disruption of normal activities by both patients and experts included abdominal pain and discomfort, bloody diarrhoea (also diarrhoea and watery diarrhoea), tiredness, loss of appetite, dizziness and nausea. There was a considerable amount of agreement in opinions between patients and experts on how symptoms may affect patients' HRQL. According to the experts, students fail to attend school, may have poor concentration and hence poor performance in school. This was corroborated by two school going patients who said, *'I was feeling weak, tired and could not do anything such as washing, eating, I could not even go to school. I was always in bed'.*

For the adults, the experts noted that they may fail to attend work and even when they do, they work less. This affects their ability to earn wages which can lead to poverty. Farmers work less hours and this may affect their productivity. There is also wastage of time, going to hospital to seek medical care, resting due to always feeling tired and frequenting toilets due to diarrhoea. Schistosomiasis may also interfere with mobility especially in chronic cases as one feels weak, is in pain and has diarrhoea. Loss of appetite also contributes to feeling weak and having no energy. These views can be summed up in some patients' reports as the following quotes show. *"I cannot go to the farm. I can't walk, cook or eat. I just feel like lying down. If I go to Ngurubani (the local market) I cannot do anything the following day. I am unable to eat. I only take two spoonfuls and I cannot eat any more. I have no happiness in doing anything"*, elderly woman. A young man had this to say, *"I feel tired like one who has worked while I have not. I cajole myself to work, but my body does not feel like. I can only dig a little and then I cannot go on. When trying to eat after two spoons, I cannot eat any more. I cannot accomplish much.....I can only work in the morning and when it gets sunny, I have to go home because I feel dizzy"*. Hence symptoms

affect people's work and work productivity, their mobility, and feelings of strength in body due to inability to eat.

Opinions on HRQL Dimensions

The majority of patients and experts seemed to agree on opinions that the eight HRQL domains were affected by bilharzia.

Mobility was affected in the sense that due to symptoms like abdominal pain and discomfort, diarrhea and tiredness patients are unable to move about. For example patients reported that when having symptoms, they can '*walk only a little... cannot go anywhere*' (three males 22, 29 and 30 years), '*... and just want to sleep*' (female 16 years).

Activities of daily living were restricted due to symptoms like feeling weak and tired, abdominal pain and discomfort, loss of appetite and bloody diarrhoea. According to experts, the presence of these symptoms reduces performance of work due to lack of participation, be it school, farm or formal work. The patient has to stop doing their usual work to go and seek medication, which means no wages for the working and poor school performance for the children. Patients said, ' *you cannot not go to work, so there is loss of work .....cannot plan your work*' (3 males 22, 29 and 30 years), '*.....cannot even eat or do daily chores in the household*' (female 16 years).

Poor workmanship, lack of attendance to work and time wasted seeking medical assistance, fewer hours put in if one goes to work leads to inadequate yields, less work production and reduced school performance according to the experts. Patients expressed similar views, '*...you work a little and get tired*' (male 22 years), '*... even though you have the willingness, the body refuses to work, you just want to rest,* (female 60 years), '*... you feel as if all the joints are dead and to wake up you need support. I have been absent from school, so I know my learning will change*' (female 16 years).

Experts noted that due to symptoms like tiredness, loss of appetite, bloody diarrhoea, the body feels weak and tired and has no energy. This was in agreement with patients. They reported feeling weak, felt they had lost weight and could not do anything.

Regarding social participation experts noted that the patient would be unable to attend due to symptoms as they limit ones movements. Patients noted that even if one attends social functions they will not concentrate and that due to the symptoms one cannot walk for long. A 16 year-old girl summed it up as, '*I just stayed home, I could not even play.....feeling sad like someone who is mourning*'.

Feelings of tension, hopelessness and despair arise due to poverty resulting from not working and poor school performance, feeling weak and unhealthy, not knowing what is wrong and as a result of swollen legs, feet and belly, according to experts. Patients attributed such feelings to having to depend on others, being unable to do their duties and due to lack of happiness and accomplishment.

Patients and experts agreed in opinion about worry and anxiety. They noted that worry and anxiety results from disfigurement (swollen belly), seeing blood in stool which is scary as patients may think the blood supply in the body will be depleted, and also due to feeling weak. According to experts, where schistosomiasis has reached advanced stages, patients worry due to severe symptoms and complications from the disease.

Thoughts about infection include people thinking they are bewitched, wondering if they will ever get cured especially with re-infection, not knowing what the infection is and also thoughts that the disease is incurable according to experts. One patient said, '*I had bad thoughts. I thought I was dying*' (female 16 years).

From these opinions and views, it is apparent that all the selected HRQL dimensions are affected in some way as a result of Schistosomiasis Mansoni infection. It is therefore appropriate to include them in the final questionnaire. However as noted earlier the last

three domains are regarded as similar feelings and worry and anxiety was used rather than all three.

5.3.4.4 Response categories for symptoms and HRQL dimensions

For each of the symptoms, there was a question assessing frequency, intensity and duration of activity disruption. Similarly, for the HRQL dimensions there was a question assessing the extent to which they were affected.

Response categories for frequency were 'never', 'very rarely', 'rarely', 'often', 'very often' and 'always'. During the pre-test, respondents were asked to state what time frame they thought of in referring to these terms. Respondents had a lot of difficulty assigning a time frame to these terms and there were wide variations in what the terms meant to different people. For example 'very rarely' was thought of in time ranging from minutes to a week by patients and from 10-minute interval in a day to once a month. These variations were observed with the other terms. However, despite the variations, experts' conceptions of these terms seemed to reflect increasing frequency. For example 'very rarely' coincided with anything less than a day, 'rarely' between a day to about three days and 'very often' anything between one to two weeks. Due to these variations in conceptions a decision was made to represent frequency in terms of number of days in the last two weeks to shorten the recall period. This would make quantifying frequency explicit and to have similar meaning for different people.

Duration of activity restriction was presented in terms of days. Since there was no problem responding to this question, the response categories were retained but made consistent with the 'in the last two weeks' recall duration adopted in the questionnaire.

Response categories for intensity of symptoms were 'mild', 'moderate', 'severe' and 'very severe'. Respondents were asked to state what each of the term meant to them. Experts described these response categories in terms of how noticeable and bothersome the symptoms were and whether they required medical attention. There was a general view of the increasing noticeability, bother and necessity to seek medical attention as intensity

increased from mild to very severe. However, problems were noted regarding description of intensity for some symptoms in these terms. Experts suggested that intensity of diarrhoea be expressed in terms of number of times stool passed per day to be consistent with definition of diarrhoea. Patients described loss of appetite in terms of amount of food they could eat. This was adapted to express loss of appetite in the final questionnaire. Tiredness was expressed in terms of 'a little, somewhat, very and extremely tired' the rest of the symptoms were expressed in terms of 'mild, moderate, severe and very severe'. Therefore response categories were changed to make sense in describing symptoms and to reflect how local people describe such symptoms.

Response categories for extent of effect on HRQL domains were 'a little of the time', 'some of the time' and 'most of the time'. These terms were described in terms of hours to days by patients and there was wide variation in the conceptions. For the experts, the description was also in terms of days but there was more consistency in terms used to describe the extent of effect on HRQL domains. However, for both patients and experts, the conception of duration increased as they moved from a little of the time to most of the time. Therefore, these response categories were retained, but they should be used with caution, as they may not mean the same thing to everyone. The decision to retain these response categories was pragmatic to avoid complications during scenario construction in the valuation phase of the study.

5.3.4.5 Assessing appropriateness of items in the questionnaire.

Assessment of opinions on suitability of items in the questionnaire to young boys and girls, young men and women as well as old men and women indicated that none of the questions were considered unsuitable, sensitive, rude or offensive.

5.3.5 Final Schistosomiasis Mansoni HRQL Questionnaire

This section describes the final questionnaire that was administered to the community and patient samples in measuring the impact of Schistosomiasis Mansoni in the empirical study. The questionnaire (see appendix 5.3)[43] has five sections. The first section seeks information on history of illness currently and in the last two weeks and subsequent activity restriction. The second section seeks information on the eleven symptoms identified in the literature reviews and patient and expert pre-test. These are abdominal pain and discomfort, diarrhoea, watery diarrhoea, bloody diarrhoea, tiredness, nausea, vomiting blood, loss of appetite, itching skin rash, fever and dizziness. For each symptom information on frequency (five response categories), intensity (four response categories), and disruption of activities and duration of disruption (five response categories) is collected. The third section collects information on whether and how six HRQL dimensions are affected due to infection with bilharzia. The fourth section contains evaluative questions to find out the meanings attached to some key concepts assessed in the questionnaire such as illness, health problems, daily duties and health state. The fifth section seeks socio-economic and demographic information of the respondent.

The patient questionnaire in addition contained an annex, which sought information on general and clinical complaints as assessed by the clinician as well as a section for recording the parasitological information on intensity of infection after conducting laboratory examination. Parasitological information was recorded in terms of eggs per gram of faeces (epgf).

Use of this questionnaire amongst patients and community members in Mwea and subsequent testing of its validity are reported in the next chapter.

[43] See attached questionnaire.

CHAPTER SIX

THE DISEASE IMPACT OF SCHISTOSOMIASIS MANSONI IN KENYA: VALIDITY OF A NEW DISEASE SPECIFIC MEASUREMENT TOOL

6.1 INTRODUCTION

In chapter five a questionnaire was constructed to assess the impact of Schistosomiasis Mansoni infection on HRQL and facilitate creation of disease state scenarios for valuation. In this chapter the results from its use amongst community members and patients with Schistosomiasis Mansoni are presented. The chapter aims to establish the validity of measurement of disease specific states amongst the Kikuyu in Kenya and to assess Schistosomiasis Mansoni disease impact on HRQL amongst people of Mwea, Kenya. It also establishes the relationships between symptoms (in terms of frequency, intensity and severity), HRQL indicators (in terms of six HRQL domains, frequency of disruption of daily duties by symptoms, a health status index and VAS rating of current health state) and infection intensity (in terms of eggs per gram of faeces (epgf)) using the patient sample. By relating these three broad groups of variables, three ways of measuring outcomes of suffering Schistosomiasis Mansoni are compared to allow exploration of construct validity of the questionnaire.

Section 6.2 presents the methods and section 6.3 reports the socio-economic, demographic, illness and health characteristics of both patients and community members. Section 6.4 presents findings on the impact of schistosomiasis on symptoms, disruption of daily duties and HRQL. In section 6.5 results of construct validity of the schistosomiasis HRQL questionnaire are presented while section 6.6 presents results of its reliability (internal consistency). Section 6.7 presents the discussion and conclusions of the results.

6.2 METHODS

6.2.1 Samples and sample selection

The schistosomiasis HRQL questionnaire was used in two sub-samples; a community and a patient sample.

Community sample: The community sample was drawn form Thiba Location, which comprises three sub-locations. Each sub-location comprises of several villages of different sizes (in terms of households), headed by a village In-Charge. Each sub-chief provided a list of all the villages and number of households per village for each of the three sub-locations. Using this information an average sized[44] village was selected in each sub-location as the study site. The three villages were Nyaikungu (Nguka sub-location), Thiba south (Thiba sub-location) and W2 (Wamumu sub-location). They had 156, 151 and 116 households respectively. The next step was to create a sampling frame upon which a random sample of households in each village would be based. With the help of the Village In-charge a household mapping of each village was done and each household in the village was assigned a number and the name of the household head obtained. At this stage, it was discovered that the number of households where there were permanent residents was 133, 113 and 116 in Nyaikungu, Thiba South and W2 respectively. Probability proportionate to size[45] was used to determine the number of households to be sampled in each village.

The required sample size of 80 respondents was calculated using the EPI INFO sample size calculation procedure for a population survey[46]. Using the probability proportionate to size and accounting for non-response and or refusals, 45, 37 and 38 households from

[44] The average size was computed by summing up all the households and dividing by number of villages per sub-location. Then the village closest in size to the average was selected as a representative village for the sub-location.

[45] This involves totalling all the households and determining the percentage contribution of each village. This percentage is then used to compute the number of households from each village for the required sample size.

[46] The EPI INFO 6 sample size calculation for a population survey asks for expected frequency plus or minus best and worst acceptable frequency and population size. With these parameters the software computes the required sample size for the 80%, 90%, 95% and 99% confidence level and selected level of power. The parameters used in the computation of the sample size were, a population size of 23,707 for Thiba (GOK, 1989), Schistosomiasis Mansoni prevalence of 20-40% amongst adults (Gryseels, 1989; Kirigia, 1998), a confidence level of 95% and 80% power, which yielded a sample size of 80.

Nyaikungu, Thiba South and W2 respectively were randomly selected (using MS EXCEL) from each village (n=120). The final step was to randomly select the respondent[47]. Any person aged 15 years and above present at the time of the visit was requested for an interview or an appointment if not available then. In cases where there was no one in the selected household, subsequent visits were made and appointments made where necessary. Therefore, the sample was random at the household level but not random at the individual level and there is therefore a potential for sample selection bias.

Patient sample: Patient recruitment was non-random and was carried out in two GOK facilities namely Kimbimbi health center and Nguka dispensary. The health center serves the whole of Mwea Division while Nguka dispensary serves Thiba Location. The two facilities had laboratory facilities, which were required for screening of patients with Schistosomiasis Mansoni infection for recruitment into the study. As in the community survey, the sample size for the patients was 80 respondents aged 15 years and above. Consecutive patients were recruited into the study, with the help of the clinical officer, laboratory technician and research assistant.

6.2.2 The Schistosomiasis Mansoni HRQL questionnaire

The contents of this questionnaire were described in section 5.4.3[48]. This section provides details of how symptoms, HRQL and infection intensity were assessed, including computation indices.

[47] Initially, this was to be accomplished by listing the names of all people in the household aged 15 years and above upon gaining consent and permission to interview. Using a random numbers card prepared for different household size numbers, a person was to be selected based on the household members' list. This was to ensure that there was no bias in selecting the respondent. However, this proved impossible in Mwea due to prevailing rice farming politics and suspicions concerning disquieting religious movements. Listing of names of those above 15 years was linked to a thorny land issue in Mwea, where those over 18 should not live in the settlement scheme. It was also feared that in the face of the violent and politically charged break up of farmers from previous control by National Irrigation Board and the ensuing confusion about rice marketing, listing of names was for the purpose of secretly tricking farmers into a formation of a rival rice marketing society against their will. Others feared that, those listed would be "sold" to devil worshipping cults. Due to these formidable problems and the uncompromising stance taken even after explaining the purpose of random sampling, the listing of names was dropped.

[48] See full questionnaire in appendix 5.3.

6.2.2.1 Symptoms

Frequency of symptoms and disruption of daily duties was assessed in terms of number of days in the last two weeks. The five frequency categories were: 1=less than a day; 2=1-3 days; 3=4-6 days; 4=7-10 days and 5=11-14 days. Intensity was assessed in four categories and differed by symptoms. However, the intensity was indicative of: 1=mild; 2=moderate; 3=severe and 4=very severe intensity. Using information on frequency and intensity, individual symptom severity, overall frequency, intensity and severity indices were computed as follows.

Individual symptom severity indices: These were computed by summing the frequency and intensity response levels for each symptom (theoretical range = 0-9, actual range = 2-9). These were then recoded to correspond to four severity levels namely, mild (sum=2 or 3), moderate (sum=4 or 5), severe (sum =6 or 7) and very severe (sum=8 or 9). Zeros were recoded as no symptoms. The severity index describes how intense and frequent a symptom was. Hence, mild symptom severity corresponds to between mild intensity of a symptom for up to 3 days and moderate intensity for less than a day. Moderate symptom severity corresponds to between mild intensity of a symptom for up to 10 days through moderate intensity for 4-6 days to very severe intensity for less than a day. Severe symptom severity corresponds to between mild intensity of a symptom for up to 14 days through moderate intensity up to 14 days to very severe intensity for 4-6 days. Finally, very severe symptom severity corresponds to between severe intensity of a symptom for up to 14 days and very severe intensity for up to 14 days.

Overall symptom severity indices: Three overall symptom indices were computed in order to gain an overall picture of symptom experience in terms of frequency, intensity and severity. To compute the overall frequency index, frequency response levels on all symptoms experienced by a respondent were summed and divided by the number of symptoms that person had. This was recoded to correspond to the five response categories for frequency. The aim of computing this index was to give an overall indication of how long on average symptoms lasted during the previous two weeks. This index therefore shows the most common duration of symptom experience. Computation of the overall intensity index followed the same procedure as frequency

index. This index provides an overall indication of the level of intensity of symptoms experienced by the respondents. The overall severity index was computed by summing the severity indices for all symptoms and dividing the sum by the number of symptoms, before recoding into similar categories as the severity index described above. This index provides an overall indication of the level of severity of symptoms.

6.2.2.2 HRQL domains

The frequency with which the six HRQL domains were affected was assessed on three levels: 1= a little of the time, 2= some of the time and 3= most of the time. This information was used to compute a health status index by summing the frequency levels reported on six HRQL domains (range 0-18) and then classifying the index into mild (1-6), moderate (7-12) and severe (13-18) health states. Zeros were classified as perfect health. Frequency of disruption of daily duties was assessed on same levels as frequency of symptoms. An overall index on frequency of disruption of daily was therefore computed following the same procedure as for the frequency of symptoms described above. This index provides an overall indication of the duration daily duties were disrupted.

6.2.2.3 Infection intensity

Intensity of infection with Schistosomiasis Mansoni was measured using number of eggs per gram of faeces. Parasitological assessment to confirm Schistosomiasis Mansoni infection status was performed by laboratory technicians based at Kimbimbi health center and Nguka dispensary using the Kato technique. Infection intensity was classified as light, moderate and heavy following WHO (1993).

6.2.3 Analytical methods

6.2.3.1 Assessment of disease impact

Data on socio-economic, demographic, illness and health problems experienced by respondents, symptoms and HRQL domains was categorical (ordered and unordered) in nature. Analysis involved reports of proportions as well as comparisons of proportions between patient and community groups. The purpose of the community group was to

provide comparisons on aspects such as illness and health problems, symptoms and HRQL domains that would inform on whether differences could be attributed to presence of Schistosomiasis Mansoni amongst patients. Where the data type was binomial i.e. two possible outcomes (e.g. yes, no) comparisons of patients and community groups (unpaired) used the Fisher's test (chi-square for large samples) while the Mann-Whitney test was used where the outcome was more than two categories (e.g. marital status or education level) to test for significant differences between proportions (Motulsky, 1995). In the case of continuous data, the t-test was used.

6.2.3.2 Assessment of construct validity

Testing of construct validity was based on responses from the patient sample because their disease status was established through laboratory testing of intensity of schistosoma mansoni ova in patients' stool samples. Construct validity was assessed using three ways of measuring Schistosomiasis Mansoni disease. These included infection intensity using eggs per gram of faeces. The second measure was impact of schistosomiasis, which comprised of six HRQL domains affected by schistosomiasis infection, disruption of daily duties by each of the symptoms and a health status index. The third measure comprised frequency, intensity and severity of symptoms as well as aggregate symptom frequency, intensity and severity indices.

In testing for construct validity the HRQL indicators, infection intensity and symptoms were assumed to be measuring the same latent variable, namely schistosomiasis disease states. It was therefore hypothesized that the worse off the health related quality of life indicators, the worse off the symptom severity and the higher the infection intensity. Similarly, the higher the infection intensity the worse off the symptom severity index.

Correlations between the three measures were used to assess construct validity. Spearman's correlation coefficient was used where the two variables were ordinal and the Cramer's V correlation coefficient was used where the two variables were nominal by nominal or ordinal (Pett, 1997). Examining the sign and magnitude of correlations establishes if there is validity. While the closer to 1 the correlation coefficient is the more validity is established, there is no recommended critical level for claims of validity. Considering the clinimetric approach adopted in constructing the questionnaire

and the causal nature of the variables, even very low correlations can still be deemed to support validity (Fayers and Hand, 2002). Positive correlations were expected between these three variables.

Further assessment of construct validity was through hypothesis testing[49]. In addition to the sign on the Spearman's correlation coefficient that establishes if the constructs are related as expected, Kruskal Wallis one-way analysis of variance was used to test the null hypothesis that different categories of respondents (as specified by grouping variables) were similar. The null hypothesis that the groups were similar was rejected in favour of the alternative hypothesis that the groups were different for $p<0.05$ (Pett, 1997; Motulsky, 1995).

Further evidence for construct validity was assessed through regression analysis by estimating the relationship between HRQL indicators on the one hand and symptoms, infection intensity and characteristics of the respondents on the other hand. Considering that the dependent variable is an ordered categorical variable (0=perfect health; 1=mild; 2=moderate and 3=severe health status), the ordered probit model is the appropriate econometric model to use (Jones, 2001: p.34). It models a discrete dependent variable that takes ordered multinomial outcomes, for example different categories of health states ordered form the best to the worst. The functional form the model is represented (Borooah, 2002) as:

$$Y_i = \sum_{k=1}^{K} \beta_k X_{ik} + \varepsilon_i$$

Where, Y_i is the dependent (explained) variable i.e. health status index and the five HRQL domains; β_k is the coefficient associated with the k^{th} variable (X) and ε_i is an error term that captures either factors left out of the equation or inaccurate measurement (Borooah, 2002). The coefficients on the explanatory variables have a qualitative interpretation (Jones, 2001). An increase in the value of the explanatory variable (X_{ik}'s) for a particular person will cause her or his explained variable Y_i to rise if β_k is positive and to fall if it is negative. Signs on coefficients were used in reporting construct validity. The odds ratios are also reported. They show the ratio, given a one-unit

[49] See appendix 6.3 for hypotheses tested.

increase in the variable, of the odds of being in a higher rather than a lower category e.g. the odds of moving from mild to above moderate or from below moderate to severe health status indicator.

The RESET (regression equation specification error test), to establish that the model specification had no omitted variables, no heteroskedasticity and non-linearity, complemented by the Ramsey RESET for omitted variables and Cook-Weisberg test for heteroskedasticity were carried out. In addition, the model's chi square value and its' level of significance as well as the pseudo-R2 were used in ensuring the fitness of the model.

6.2.3.3 Assessment of reliability (Internal consistency)

Owing to time and financial limitations, test-retest reliability of the measurement tool was not assessed. However, internal consistency that requires administering the tool once was assessed. Internal consistency or scale reliability is the extent to which the items comprising a scale measure the same concept. This is assessed using the KR-20 *(Kuder-Richardson formula 20*) for items assessed dichotomously and Cronbach's alpha coefficient when the number of response categories exceeds two (Streiner and Norman, 1995; Cano, 2001). Higher values indicate greater internal consistency, although (Streiner and Norman, 1995) caution against uncritically accepting higher values because alpha could be increased by increasing the items comprising a scale. It has been recommended that alpha coefficients be greater than 0.7 when comparing different groups because the scale may be tapping more than one attribute and should be 0.9 for individual assessment (Nunnally and Bernstein, 1994) where items are expected to be homogenous. However, if the alpha is too high this may be an indication of item redundancy, in which some of the items in the scale are asking the same question in different ways and therefore alpha should not exceed 0.9. It has been argued that item-total correlations of less than 0.2 should be discarded (Streiner and Norman, 1995) because when correlation between a single item and its scale is low, the item is probably measuring something different from other items in the scale. However, this could be expected to be lower for a scale comprising causal variables (Fayers and Hand, 2002).

Assessing internal reliability in this thesis was based on the patient sample since the measure was developed for use with this group. Items comprising frequency, intensity and frequency of disruption of daily duties for symptoms and HRQL domains were assessed as symptoms and HRQL sub-scales and then combined and assessed as total scale.

6.3 CHARACTERISTICS OF STUDY SUBJECTS

6.3.1 Socio-economic and demographic characteristics

Table A6.1 in appendix 6.2 shows that the majority of respondents in both samples were females, married, with primary level education and farmers. However, there was a higher proportion of businesspersons and students in the patient sample and fewer patients had secondary level education. Relative to the community sample, patients were significantly younger ($p<0.001$). Both samples were largely similar in terms of gender and education level while they differed significantly in terms of age, marital status and occupation.

6.3.2 Illnesses and health problems

Table A6.2 in appendix 6.2 shows that there were no significant differences in proportions of patients and community members reporting illness and being aware of what illness they had in the two weeks prior to the interviews. However, during the interviews, there were significantly more patients reporting illness and being aware of what illness they had than the community members. It was also evident that the proportion reporting various disease problems in the previous two weeks was fairly similar in the two groups while this differed significantly during the interviews, with 71.3% of patients reporting Schistosomiasis Mansoni and Schistosomiasis Mansoni related symptoms. Although awareness[50] of Schistosomiasis Mansoni amongst patients and community members was high, significantly more patients had suffered from

[50] Awareness here refers to having ever suffered and having known of anyone suffering from Schistosomiasis Mansoni.

Schistosomiasis Mansoni and known someone suffering form Schistosomiasis Mansoni than community members.

6.4 IMPACT OF SCHISTOSOMIASIS

6.4.1 Symptom experience and bother

Table 6.1 below shows the proportion of patients and community members reporting presence of symptoms in decreasing magnitude as well as those symptoms that bothered them most. The most commonly reported symptoms in both groups were abdominal pain and discomfort and tiredness and the least reported was diarrhoea. Vomiting was hardly mentioned in either case and will not be considered further. However, there was strong evidence that a higher proportion of patients than community members reported presence of all the symptoms. The largest differences between the two groups ranged between 42.8% and 55.4% for loss of appetite, fever, abdominal pain and discomfort, nausea, tiredness and dizziness.

Table 6.1: Symptom reporting and bother (% frequency) [community, n=81; patients, n=80]

symptoms	% reporting symptom presence			Symptoms bothering most (%)		
	Patients	Community	difference in proportions (P -C)	Patients	Community	difference in proportions ▪ (P -C)
Abdominal pain and discomfort	91.3	43.2	48.1**	53.8	25.9	27.9
Tiredness	91.3	46.9	44.4**	11.3	11.1	0.2
Fever	87.5	32.1	55.4**	2.5	1.2	1.3
Nausea	80.0	32.1	47.9**	1.3	4.9	-3.6
Loss of appetite	80.0	24.7	55.3**	2.5	1.2	1.3
Dizziness	66.3	23.5	42.8**	16.3	6.2	10.1
Itching skin rash	38.8	14.8	24.0**	2.5	1.2	1.3
Bloody mucoid diarrhoea	37.5	7.4	30.1**	5.0	2.5	2.5
Watery diarrhoea	25.0	1.2	23.8**	1.3	1.2	0.1
Diarrhoea	22.5	6.2	16.3**	1.3	1.2	0.1
Vomiting	1.3	0.0	1.3	-	-	

Test statistic, Chi-square Fisher's exact test. *P values* - ** <0.01. ▪ test for significance used Mann Whitney test and $p<0.001$.

Both groups reported that abdominal pain and discomfort, tiredness and dizziness were the symptoms that bothered them most. Bloody mucoid diarrhoea and nausea were mentioned by four patients and community members respectively, while all the other symptoms were mentioned only once or twice. The Mann-Whitney test of differences between the two groups showed that significantly more patients reported being bothered by the symptoms than the community members.

Table A6.3 in appendix 6.2 presents the frequency, intensity and severity of symptoms for patients and community members and differences between proportions experiencing a symptom on different levels. *P* values based on Mann-Whitney test for differences between these proportions are also presented. Frequency, intensity and severity of symptoms are reported separately below.

Frequency of symptoms: In general between 21% and 38% of the patients experienced symptoms such as abdominal pain and discomfort, bloody diarrhoea, tiredness, nausea, loss of appetite, fever and dizziness for periods ranging between 1-6 days while for community group only abdominal pain and discomfort and tiredness were experienced for 1-3 days by over 20% of respondents. Symptoms were experienced for between less than a day to two weeks for the commonly occurring symptoms for both patient and community groups. However, the majority of frequency categories were reported by less than 10% of respondents. In general most symptoms lasted for about one week with a few respondents experiencing both shorter and longer durations (figure A6.1 in appendix 6.1). In terms of differences in proportion of patients and community members reporting a symptom for any given duration, there was strong evidence that significantly more patients than community members experienced all these symptoms.

Intensity of symptoms: 20-50% of patients reported moderate to severe intensity for abdominal pain and discomfort, tiredness, nausea, loss of appetite, fever and dizziness, while a similar proportion reported a mild intensity on diarrhoea and bloody diarrhoea. 10-20% of community members reported a mild intensity for nausea, moderate intensity on loss of appetite and severe intensity on tiredness and fever. Reports of symptom intensity ranged from mild to very severe for all symptoms except watery diarrhoea and diarrhoea amongst patients, which was similar in community except for loss of appetite where intensity ranged from moderate to very severe. With the exception of bloody diarrhoea, watery diarrhoea and diarrhoea, where mild intensity was more common, all other symptoms tended to occur with moderate to severe intensity with mild and very severe intensity being experienced less often (figure A6.2 in appendix 6.1). In general, there was strong evidence of significantly more patients than community members reporting all symptoms on intensity levels.

Individual Symptom severity indices: Figure A6.3 in appendix 6.1 shows that in the patient group, between 30% and 46% experienced moderate severity while 24% to 41% experienced severe abdominal pain and discomfort, tiredness, nausea, loss of appetite, fever and dizziness. In the community group, between 10% and 19% experienced moderate severity while 7% to 13% experienced severe symptoms similar to those in patients except dizziness. Symptom severity amongst patients ranged from mild to very severe for all symptoms, except watery diarrhoea and diarrhoea where severe and very

severe intensity was not experienced. The trend was similar amongst community except for itching skin rash and bloody diarrhoea. The proportion of patients experiencing various symptom severity levels was significantly higher than for the community. The evidence was extremely strong for all symptoms.

Overall symptom severity indices: Distributions of respondents across overall frequency, intensity and severity levels are reported in figure A6.4 (in appendix 6.1). The most common duration of symptoms, based on the aggregate frequency index, was 1-6 days with 86.3% of patients falling in this category compared to 46.9% in the community group. In terms of intensity of symptoms, the common category was moderate intensity experienced by 75% of patients compared to 38.3% of the community group. Finally, common symptom severity was moderate to severe accounting for 95.1% among patients compared to 55.5% in the community group.

6.4.2 Impact of schistosomiasis on HRQL: Disruption of daily duties and HRQL domains

6.4.2.1 Disruption of daily duties

Table A6.4 in appendix 6.2 shows that abdominal pain and discomfort, tiredness, nausea, loss of appetite, fever and dizziness were reported as disrupting daily duties by between 57.5% to 80% by patients compared to 17.3% to 35.9% in community group. A significantly higher proportion of patients than community members reported that all symptoms disrupted daily duties. The largest differences occurred in reports of abdominal pain and discomfort, loss of appetite, nausea and dizziness while the least was in diarrhoea. Types of daily duties disrupted included usual work (farming, casual labour, business), household chores (cooking, fetching water, cleaning homestead and personal hygiene), mobility (being able to go to places) and schoolwork.

Figure A6.5 in appendix 6.1 shows that all symptoms were reported as disrupting daily duties for varying duration. Duration of disruption of daily duties by all symptoms except diarrhoea ranged from less than a day to two weeks amongst patients. In the community groups, duration of disruption of daily duties by all symptoms except itching skin rash, bloody diarrhoea, watery diarrhoea and diarrhoea, ranged from less

than day to two weeks. Amongst the patients, between 20% and 37.5 % reported abdominal pain and discomfort, tiredness, nausea, loss of appetite, fever and dizziness as disrupting daily duties for between 1-6 days. In the community group, abdominal pain and discomfort, tiredness and fever were reported by between 11.1% and 17.3 % as disrupting daily duties for 1-3 days and between one to two weeks for tiredness. The Chi-square Fisher's exact test of difference between proportion showed strong evidence of higher proportions of patients reporting disruption of daily duties than did community members (table A6.4 in appendix 6.2).

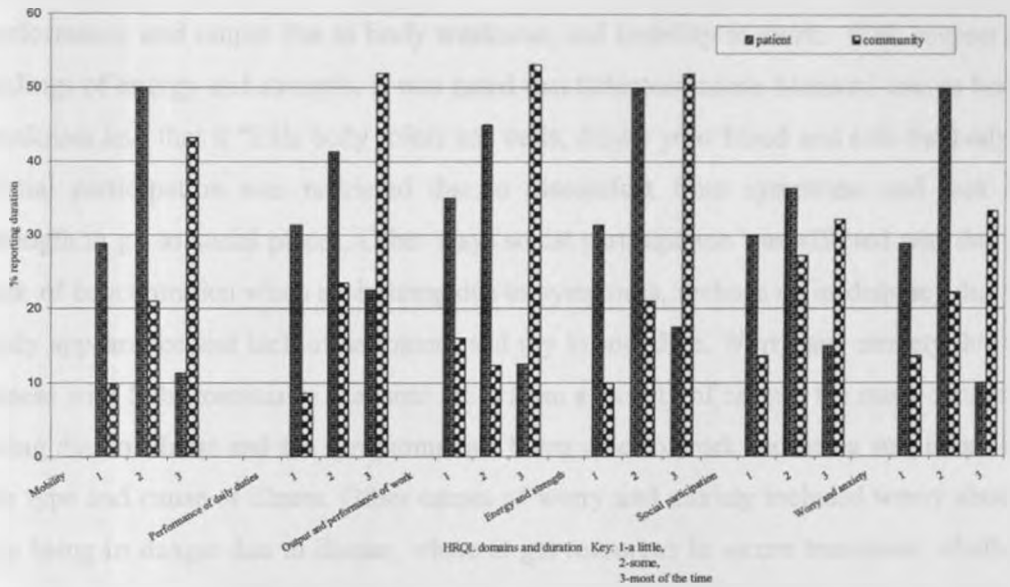
6.4.2.2 Impact on HRQL Domains

Respondents were asked to indicate agreement or disagreement regarding whether being ill with Schistosomiasis Mansoni would affect the six HRQL domains. They were also asked to state how often the domains would be affected and how they would be affected. Figure A6.6 in appendix 6.1 shows that the majority (over 70%) of both patient and community members stated that being ill with Schistosomiasis Mansoni would have an impact on all the six HRQL domains, with no significant[51] differences between the two groups. In both groups, mobility, performance of daily duties, output and performance of work and feeling of energy and strength in body had higher proportions of respondents stating that they would be affected compared to social participation and feelings of worry and anxiety.

For all the domains, statistically significantly[52] more patients (range: 28.8%-35% and 36.3%-50%) than community members (range: 8.6%-16% and 12.3% to 27.2%) indicated that the domains would be affected a little of the time and some of the time respectively. On the contrary, statistically significantly more respondents from the community group (32.1%-53.1%) compared to patient group (10%-22.5%) indicated that the domains would be affected most of the time (figure 6.1 below).

[51] Used chi-square Fisher's exact test for differences in proportion between those agreeing and disagreeing.
[52] Tested using Mann-Whitney test. All *P* values were less than 0.002.

Figure 6.1:Duration that HRQL domains would be affected (%). [all p<0.002]



Respondents were asked to describe ways in which being ill with Schistosomiasis Mansoni would affect the six HRQL domains if they agreed that they would be affected. Since it was not known whether community members had Schistosomiasis Mansoni or not at the time of the interviews, they were asked to state how they thought Schistosomiasis Mansoni would affect the HRQL domains based on their experience with Schistosomiasis Mansoni either personally previously or through a family or community member. The patients were however asked how Schistosomiasis Mansoni affected them currently due to their illness with Schistosomiasis Mansoni. This will become important in analysing the relationships and testing for validity in that only patient responses will be used, given that their Schistosomiasis Mansoni status was known.

Considering that responses from the community members cannot be directly attributed to illness with Schistosomiasis Mansoni, they are excluded from reports of how the six HRQL domains were affected although their hypothetical responses are included in table A6.5 (appendix 6.2) alongside those of the patients. Restriction in mobility was attributed to severity of symptoms resulting in inability to walk or move about. Performance of daily duties was affected due to inability to work as a result of the symptoms and body weakness, lack of progress in work due to inability to go to work,

and reduction in working hours if and when one attended. It was also noted that Schistosomiasis Mansoni affected performance and output of work through reduction in performance and output due to body weakness, and inability to work. With respect to feelings of energy and strength, it was noted that Schistosomiasis Mansoni causes body weakness and that it "kills body joints and cells, drinks your blood and eats the body". Social participation was restricted due to discomfort from symptoms and lack of strength to go to social places. Other ways social participation was affected was due to lack of concentration when socializing due to symptoms, feelings of inadequacy due to body appearance and lack of happiness and joy to socialize. Worry and anxiety due to illness with Schistosomiasis Mansoni arose from a variety of causes, the most common being due to illness and the symptoms, not being able to work and being suspicious of the type and cause of illness. Other causes of worry and anxiety included worry about: life being in danger due to illness; where to get resources to secure treatment; whether one will get cured; likelihood of dying; lack of happiness and restlessness of the mind.

In general Schistosomiasis Mansoni was considered to affect all the six HRQL domains in various ways. Most of the respondents concurred that Schistosomiasis Mansoni had adverse effects on each of the domains and also indicated how these domains would be affected as a result of being ill with Schistosomiasis Mansoni.

6.4.3 Infection intensity

Table 6.2 below shows intensity of infection with Schistosomiasis Mansoni amongst the patients sampled, measured using number of eggs per gram of faeces. The higher the concentration of eggs per gram of faeces the heavier the infection intensity. The mean infection intensity was 352 epgf (eggs per gram of faeces) with a standard deviation of 272 epgf, suggesting moderate intensity. The majority (48.8%) of the patients had moderate infection intensity.

Table 6.2: Infection intensity (eggs per gram of faeces) n=80

Infection intensity (% of patients)	
Light (<100)	18.8
Moderate (100-400)	48.8
Heavy (>400)	32.5
Mean intensity (standard deviation)	352 (272)
Inter-quartile range	150-250

6.5 CONSTRUCT VALIDITY OF THE SCHISTOSOMIASIS HRQL QUESTIONNAIRE

Symptoms were assessed in terms of frequency, intensity, severity and three overall indices based on frequency, intensity and severity. The HRQL indicators comprised of how often and in what ways six HRQL domains were affected by Schistosomiasis Mansoni, how often daily duties were disrupted by each of the symptoms and in what ways, and two overall indices based on the six HRQL domains and frequency of disruption of daily duties. Infection intensity was assessed in terms of eggs per gram of faeces.

As stated earlier construct validity was assessed using correlation analysis and tests of hypothesis. The following broad hypotheses[53] were tested.

1. Patients with higher infection intensity have worse off HRQL.

2. Patients with higher infection intensity have higher symptom severity index.

3. The more severe the symptoms the more HRQL domains are affected.

The Kruskal Wallis one way analysis of variance test, conducted amongst HRQL indicators and symptom severity using infection intensity (epgf) as the grouping variable was used to test the above hypotheses. The test reveals whether there are any differences in health status index and accompanying HRQL domains, symptoms

[53] See appendix 6.3 specific hypothesis.

severity indices (for individual symptoms and aggregated indices) between groups. Patient's VAS rating of current health state was also used for additional evidence on construct validity. Correlations between components of the three measures, together with Kruskal Wallis tests of hypotheses are reported in sections 6.5.1. to 6.5.3.

6.5.1. Association between HRQL indicators and symptoms

In this section, three sets of relationships are examined in terms of correlation and hypothesis testing. The first set correlates symptoms (frequency, intensity, severity and aggregate indices) with how often the HRQL domains are affected and the heath status index. The second set correlates symptoms with reports of ways in which HRQL domains were affected. The third set correlates symptoms with how often they disrupted daily duties. This set of correlations relates directly to the frequency with which symptoms affected daily duties thereby providing an additional check for validity.

Figures A6.7-A6.8 in appendix 6.1 and figure 6.2 below show that the majority of associations between frequency, intensity, severity of symptoms and the HRQL domains were positive, except for diarrhoea. The highest positive correlation coefficient between HRQL domains and symptom frequency, intensity and severity were 0.31, 0.28 and 0.33 respectively. Specifically, the more frequent, intense and severe the loss of appetite and itching skin rash were, the more often the HRQL domains were affected and this also held for frequency and severity of abdominal pain and discomfort, tiredness, dizziness and bloody diarrhoea. A few unexpected negative correlations were found between different symptoms and HRQL domains. For example, all aspects of fever correlated negatively with mobility and performance of output and work. A similar relation was found between all aspects of watery diarrhoea and performance of output and work and worry and anxiety. However, all the negative associations were very weak (below -0.18) and statistically insignificant. Table A6.6 in appendix 6.2 (rows 2-4) shows the HRQL domains that had positive and significant correlations with symptom frequency, intensity and severity respectively, at $p<0.05$. Tiredness, nausea, loss of appetite dizziness and itching skin rash had positive and statistically significant correlations with at least one of the six HRQL domains except performance of output

and work. Although the correlations are weak to moderate, that the majority are positive suggests construct validity between HRQL domains and symptoms.

The Kruskal Wallis test of hypothesis produced evidence of construct validity for bloody diarrhoea and loss of appetite with regard to the six HRQL domains. The more severe the bloody diarrhoea, the more often social participation was affected (p=0.043) and the more often the patient experienced worry and anxiety (p=0.009). Similarly the more severe the loss of appetite was the more often performance of daily duties (p=0.031) and social participation (p=0.012) were affected. Table A6.7 shows that all the other individual symptom severity indices failed to support the hypothesis, although figure 6.2 shows that the majority of correlations were positive.

Figure 6.2: Correlations between severity of symptoms and HRQL domains (patient group) [correlations >0.22 significant at p<0.05]

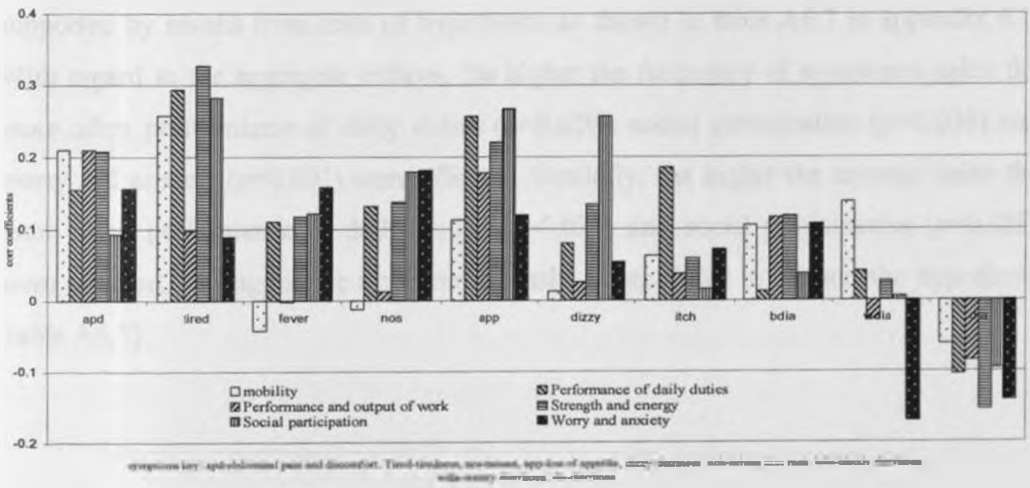
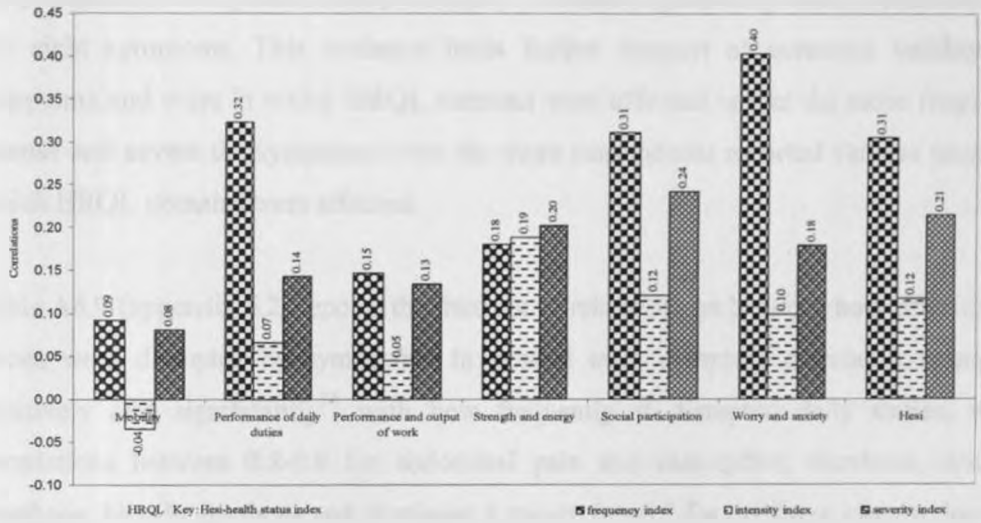


Figure A6.9 (appendix 6.1) shows that aggregate frequency, intensity and severity indices on all symptoms except diarrhoea and frequency of watery diarrhoea correlated positively with health status index. The highest correlation coefficients were around 0.25-0.31. All aspects of loss of appetite correlated positively and significantly with the health status index while only frequency and severity of abdominal pain and discomfort and tiredness had positive and significant correlations with the health status index. From tests of hypothesis, there was support of construct validity between severity of abdominal pain and discomfort (p=0.05), loss of appetite (p=0.027) and health status

index. Figure A6.10 in appendix 6.1 shows that all other symptom severity indices including the aggregate frequency, intensity and severity index failed to support the hypothesis that the more severe the symptoms the worse off the health status index.

Figure 6.3 below shows the correlations between overall indices on frequency, intensity and severity of symptoms and HRQL domains together with the health status index. With the exception of intensity index and mobility which had very weak and highly insignificant negative correlation, all other indices correlated positively with the HRQL domains and the health status index. The highest correlation coefficients for frequency, intensity and severity index were 0.40, 0.19 and 0.24 respectively. Correlations that were significant at $p<0.05$ included frequency index and performance of daily duties, social participation, worry and anxiety and the health status index as well as severity index and social participation. This overall finding supports the claim for construct validity as the more frequent, intense and severe the symptoms, the more often HRQL domains were affected and the worse off the health status index. This finding was supported by results from tests of hypotheses as shown in table A6.7 in appendix 6.2. With regard to the aggregate indices, the higher the frequency of symptoms index the more often performance of daily duties (p=0.029), social participation (p=0.036) and worry and anxiety (p=0.003) were affected. Similarly, the higher the severity index the more often performance of daily duties (p=0.027) and social participation (p=0.029) were affected. The aggregate symptom intensity index failed to support the hypothesis (table A6.7).

Figure 6.3: Correlations between aggregate symptom indices, HRQL domains and health status index (patient group) [correlations >0.24 significant at p<0.05]



VAS rating of current health was correlated with symptom severity indices as a further check for construct validity between HRQL and symptoms. Figure A6.11 in appendix 6.1 shows that there were negative correlations between VAS rating of current health and symptom severity for all symptoms except diarrhoea and itching skin rash. Abdominal pain and discomfort, bloody diarrhoea, tiredness, loss of appetite and aggregate severity index had negative and significant correlations ($p<0.05$). This evidence supports construct validity of severity of symptoms in relation to VAS rating of current health as a HRQL indicator. Tests of hypotheses that the higher the symptom severity index the lower the VAS rating of current health was supported with respect to watery diarrhoea, bloody diarrhoea and the total symptom severity index ($p<0.05$) (figure A6.12 in appendix 6.1).

Table A6.8 (appendix 6.2) presents the second set of relationships. The Cramer's V correlation coefficients between symptoms and ways in which HRQL domains were affected were all positive and ranging between 0.17-0.51, 0.19-0.52 and 0.16-0.47 for frequency, intensity and severity of symptoms respectively. The highest correlations tended to be with worry and anxiety, strength and energy and social participation, while the lowest were with mobility. Table A6.6 in appendix 6.2 (rows 7-10) shows the HRQL domains that had positive and significant correlations with symptoms at $p<0.05$.

All symptoms except dizziness and itching skin rash correlated positively and significantly with at least one of the six HRQL domains, while mobility, performance of daily duties and worry and anxiety tended to correlate significantly with at least five of the eight symptoms. This evidence lends further support of construct validity of symptoms and ways in which HRQL domains were affected in that the more frequent, intense and severe the symptoms were the more respondents reported various ways in which HRQL domains were affected.

Table A6.9 (appendix 6.2) reports the third set of relationships between how often daily duties were disrupted by symptoms. In general each symptom correlated strongly, positively and significantly[54] with how frequently it disrupted daily duties, with correlations between 0.8-0.9 for abdominal pain and discomfort, diarrhoea, watery diarrhoea, bloody diarrhoea and dizziness; between 0.6-0.7 for tiredness, nausea, loss of appetite, itching skin rash and fever. The majority of correlations for other symptoms and how frequently they disrupted daily duties were positive with the highest being 0.43, 0.37 and 0.39 for frequency, intensity and severity respectively, and only those above 0.22 were statistically significant at $p<0.05$[55]. In a few cases, unexpected negative correlations were found. These involved mainly watery diarrhoea, diarrhoea, itching skin rash and in one or two cases bloody diarrhoea and fever. Most of the negative correlations were however weak and insignificant except for those between diarrhoea, and bloody diarrhoea and loss of appetite. Table A6.10 in appendix 6.2 shows that the hypothesis that the higher the severity index for individual symptoms the higher the disruption of daily duties by individual symptoms was supported for several symptoms. All individual symptom severity indices (except bloody diarrhoea) adduced support for the hypothesis for disruption of daily duties in between one to five instances with p values ranging between 0.002 for loss of appetite to 0.049 for diarrhoea and abdominal pain and discomfort, adducing evidence for construct validity.

Figure A6.13 (appendix 6.1) shows that except for itching skin rash and diarrhoea, correlations between symptoms and the total index of disruption of daily duties were positive. All correlations above 0.22 were significant at $p<0.05$ and involved abdominal pain and discomfort, tiredness, fever, nausea, loss of appetite and dizziness. The

[54] See entries in bold in table A6.9.
[55] See entries in bold in table A6.9

hypothesis that the more severe the individual symptoms the higher the index of disruption of daily duties was supported for diarrhoea, tiredness, loss of appetite, fever and dizziness ($p<0.05$) (table A6.10 in appendix 6.2).

Figure A6.14 (appendix 6.1) shows correlations between aggregate indices based on frequency, intensity and severity of symptoms and disruption of daily duties. All indices on tiredness, fever, nausea and loss of appetite correlated positively and significantly with frequency and total index on disruption of daily duties. Only frequency and severity index on abdominal pain and discomfort and dizziness and severity index on itching skin rash and bloody diarrhoea had positive and significant correlations with frequency of disruption of daily duties. The hypothesis that the higher the aggregate frequency, intensity and severity indices the higher the frequency of disruption of daily duties was supported (table A6.10 in appendix 6.2). The higher the aggregate frequency index, the more often daily duties were disrupted due abdominal pain and discomfort, tiredness, and loss of appetite ($p<0.05$). With regard to aggregate symptom intensity index, the higher the index the more often daily duties were disrupted due to tiredness, nausea, loss of appetite and fever ($p<0.05$). Similarly, the higher the aggregate symptom severity index the more often daily duties were disrupted due to abdominal pain and discomfort, tiredness, nausea, loss of appetite, fever and dizziness ($p<0.05$). It was also found that the higher the aggregate frequency, intensity and severity indices the higher was the frequency of disruption of daily duties ($p<0.004$). This evidence supports the construct validity between symptoms and how often they disrupted daily duties.

6.5.2 Association between HRQL indicators and infection intensity

Figures A6.15 and A6.16 in appendix 6.1 show correlations between HRQL indicators and infection intensity. All correlations between HRQL domains, health status index and infection intensity were negative, weak (-0.05 to -0.16) and not statistically significant. This suggests that the heavier the worm load or infection intensity the less often HRQL domains were affected. However, correlations between infection intensity and how often daily duties were disrupted were positive except when disruption of daily duties was due to tiredness and nausea. The correlations were very low and only that between infection intensity and disruption of daily duties due to fever was significant at

$p<0.05$. This mixed evidence, most of which is in conflict with the hypothesis, is inconclusive and therefore does not lend support of construct validity between infection intensity and how HRQL domains were affected.

Due to the conflicting evidence of very low and insignificant positive and negative correlations, construct validity of infection intensity and HRQL domains remains ambiguous. However, if we focus on correlation between infection intensity and health status index (-0.09) as well as the HRQL domains, infection intensity cannot be used as an indicator of how often and how Schistosomiasis Mansoni affects HRQL domains and disruption of daily duties. These two measures therefore do not appear to be tapping the same latent variable in patients.

Figures A6.17 and A6.18 in appendix 6.1 present the *p*-values for Kruskal Wallis tests of hypothesis that the higher the infection intensity the worse off the health status. The Kruskal Wallis one way analysis of variance test for differences between categories of infection intensity and health status index did not support the hypothesis that the higher the infection intensity the worse off the health status index (p=0.105). With regard to individual HRQL domains all the p values were above 0.05 thereby lending no support for construct validity. Similarly, test of hypotheses that the higher the infection intensity the higher the reported frequency on disruption of daily duties by individual symptoms and the total index of disruption of daily duties was not supported (all p>0.153). However, in regard to VAS rating of current health, the hypothesis that the higher the infection intensity the worse off the VAS rating of current health state was supported both by the Kruskal Wallis test (p=0.043) and the negative and significant correlations (r=-0.281; p=0.012). In general, there is more evidence failing to support construct validity between infection intensity and HRQL domains, with only one piece of evidence lending support.

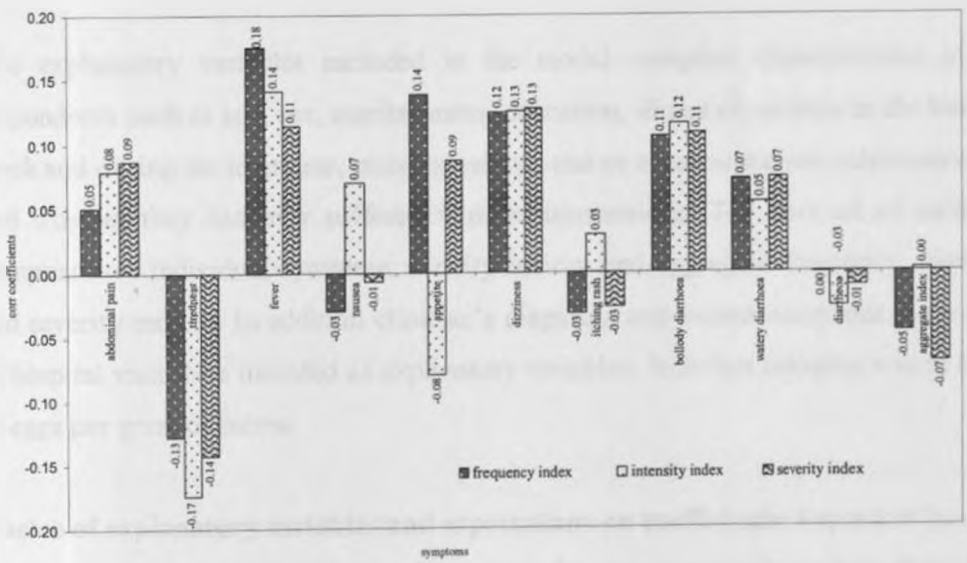
6.5.3 Association between infection intensity and symptoms

Figure 6.4 below shows that the more frequent, intense and severe abdominal pain and discomfort, fever, dizziness bloody diarrhoea and watery diarrhoea was, the more intense was the infection. Intensity of nausea and itching skin rash and frequency and

severity of loss of appetite were associated with higher infection intensity. However, frequency, intensity and severity of tiredness and diarrhoea was associated with lighter infection intensity, as was frequency and severity of nausea and itching skin rash and intensity of loss of appetite. None of the correlations were statistically significant. These findings based on individual symptoms and aggregate indices fail to support or reject the construct validity of symptoms and infection intensity.

It was hypothesised that patients with higher infection intensity would also have a higher severity index on individual symptoms and aggregate frequency, intensity and severity indices. Figure A6.19 in appendix 6.1 shows that this was not supported by the findings (p>0.273). Hence, there was no strong evidence to support construct validity between infection intensity and symptom severity.

Figure 6.4: Correlations between intensity of infection and frequency, intensity and severity of symptoms (patient group) [no correlations significant at p<0.05]



6.5.4 Does infection intensity and symptom severity explain variation in HRQL due to Schistosomiasis Mansoni?

6.5.4.1 The model and explanatory variables

To further explore the relationship between HRQL indicators and infection intensity as well as symptom severity, ordered probit regression analysis was performed. This regression technique was selected based on the ordered categorical nature of the explained variable (Jones, 2001). Both aggregated and disaggregated models were estimated with respect HRQL indicators. The aggregated model used the grouped health status index as the explained variable. The disaggregated models used mobility, performance of daily duties, feeling of energy and strength, social participation and worry and anxiety as the explained variables. Considering that tests of construct validity using correlation found performance and output of work redundant, regression analysis was not performed on this HRQL domain.

The explanatory variables included in the model comprise characteristics of the respondents such as age, sex, marital status, education, illness experience in the last two week and during the interview, prior experience and or awareness about schistosomiasis and whether they had ever suffered from schistosomiasis. The next set of variables comprised of individual symptom severity indices and aggregate frequency, intensity and severity indices. In addition clinician's diagnosis and patient complaint at the time of hospital visit were included as explanatory variables. Infection intensity was in terms of eggs per gram of faeces.

Choice of explanatory variables and expectations on coefficients: Reports of how the HRQL domains upon which the health status index was computed was hypothesized to depend on socio-economic and demographic characteristics, experience with illness, severity of illness and infection intensity.

Prevalence and intensity of infection with schistosomiasis is highly age dependent (Guyatt, 1998; Butterworth et al. 1991, 1994 and Chandiwana and Christensen, 1988) with the younger bearing more burden. However, severe disease follows after many

years (often after 20 years of age) of silent mildly symptomatic infection (WHO, 1993). Considering that the study had a lower age limit of 15 years and the delayed onset of severe morbidity, it cannot be stated a priori what sign the age coefficient would have although literature suggests that reports of health status would vary inversely with age.

There was no information directly related to schistosomiasis disease states upon which expectation on sex, marital status, education level, knowledge and experience with illness could be based. Hence the signs on these variables cannot be stated a priori.

Individual symptom severity indices as well as aggregate frequency, intensity and severity indices are expected to vary positively with health status index such that the worse off the symptom severity the worse off the health status index. Hence, coefficients on these variables are expected to be positive. Similarly, it was expected that those with higher infection intensity were experiencing more disease and bother from symptoms and were therefore likely to report worse off health status. However, results from correlation analysis do not support this and hence the coefficient on infection intensity cannot be stated a priori. Variables on clinician's diagnosis, patient's complaint at the time of hospital visit and other parasites were included to assess the likelihood of attributing variation in reports of health status to schistosomiasis. Since the hypothesis was that schistosomiasis affects health status adversely, positive coefficients were expected for clinical diagnosis, patient complaint and presence of other parasites implying more likelihood of reporting worse off health states due to illness.

6.5.4.2 Regression diagnostics

Normality

Tests of normality concern the explained variable. The Shapiro-Francia test for normality and normal data (STATA 7.0) revealed that the variable health status index, how often mobility, performance of daily duties, feeling of energy and strength, social participation and worry and anxiety were normally distributed (p=0.93 to 1.000). Of the explanatory variables, gender, marital status, education level, illness in the last two weeks, clinician's diagnosis, symptom severity indices on diarrhoea, watery diarrhoea, bloody diarrhoea, tiredness, nausea, itching skin rash and dizziness and all three aggregate severity indices were normally distributed according to Shapiro Francia test for normal data ($0.114<p\leq1.000$). The variables age, currently ill, ever suffered from

schistosomiasis, knowledge of schistosomiasis, presence of other parasites, patients complaint at hospital visit, symptom severity index of abdominal pain and discomfort, loss of appetite and fever were not normally distributed ($p \leq 0.05$). Logarithmic transformation of continuous variables, age and eggs per gram of faeces rendered them normal. The rest of the variables were retained as they were since they were categorical.

Missing data

During the interviews, care was taken to maximize response rates. The only missing data on the regression variables was due to non-applicability of the questions rather than refusal to answer, overlooking questions, not knowing or neglect. Missing data was mainly for those symptoms that the respondents did not experience and ranged between 8.75% for abdominal pain and discomfort to 77.5% for diarrhoea. These responses were coded as no symptoms. Missing data was therefore not considered as a problem.

Model selection and specification (RESET test)

To select the variables and the model, ordered probit regression was run using all variables[56]. In addition each variable was regressed individually on health status index using ordinary least squares and ordered probit models. Variables with significant *p* values ($p<0.2$) in any of the regression were all considered for the reduced model.

The regression diagnostic tests employed in choosing the model included the RESET (regression equation specification error test) F-statistic, Ramsey RESET F-statistic, Cook-Weisberg χ^2, pseudo R^2, and log likelihood χ^2. The value of pseudo R^2 increases as the fit of the model improves (Borooah. 2002). The value of χ^2 and its *p* value are indicative of the explanatory power of the model than the "intercept only" model. If the values are very large and the *p* value is significant, then this implies that the model has greater explanatory power. The RESET tests for heteroskedastic residuals which could result from heteroskedastic error terms, omitted variables as well as misspecification

[56] An OLS model was estimated using the unclassified health status index (2-18) as the explanatory variable. The RESET test for model specification and nonlinearity and Ramsey RESET test for omitted variables consistently indicated that the model was mis-specified. Since the Cook-Weisberg test for heteroskedasticity revealed that heteroskedasticity was not present, it was assumed that there was non-linearity and or some variables were omitted. Consequently, the OLS regression estimation was not pursued any further as it would result in estimators that are not BLUE (Best Linear Unbiased Estimators).

due to non-linearity. The test is based on regressing the squared residuals on the explanatory variables and their cross products and uses the overall F-test to test if the regression is significant or not. If the F-statistic is significant, the null hypothesis of homoskedasticity is rejected, thereby indicating presence of heteroskedasticity or regression mis-specification (Mukherjee et al. 1998). Heteroskedasticity can result from heteroskedastic error terms, model mis-specification due to omitted variables or non-linearity and the RESET test does not indicate which is the culprit (Mukherjee et al. 1998). Therefore, the Ramsey RESET test for omitted variables and the Cook-Weisberg test for heteroskedasticity were also performed. Significant F-statistic for Ramsey RESET and χ^2 for the Cook-Weisberg test indicate presence of heteroskedasticity and ommited variables.

6.5.4.3 Ordered probit analysis results

Variable selection resulted in a model that was specified as:

> HSI = f (sex, log of age, marital status, education level, experience of illness the previous two weeks, illness during valuation, ever suffered bilharzia, log of infection intensity, presence of other parasites, clinician's diagnosis, patient complaint, severity of symptoms, aggregate frequency, intensity and severity indices); where HSI is health status index.

All models (i.e. grouped health status index, mobility, performance of daily duties, energy and strength, social participation and worry and anxiety) with the above variables included were subjected to diagnostic tests. Table 6.3 shows that all models except performance of daily duties passed the regression diagnostic tests with aggregate severity index excluded because it was highly correlated with frequency index. The performance of daily duties model passed the regression diagnostic tests with the aggregate severity index rather than frequency index included. Hence, these models for which results are reported in table 6.3, there were no omitted variables, no heteroskedasticity and the models were correctly specified as the test statistics were all not statistically significant. Although the log likelihood $\chi 2$ was high and all p values significant indicating satisfactory explanatory power of the model, the Pseudo R^2 values were modest (0.17 for social participation model to 0.32 for the health status index model).

The next sections present the results (table 6.3) on relationship between HRQL indicators and socio-demographic, infection intensity and symptom severity variables respectively.

Table 6.3: Ordered probit regression analysis results

	Health status group model			mobility model			performance of daily duties model			energy model			social model			worry and anxiety model		
Variables	Coeff.	odds ratios	[P value]	Coeff.	odds ratios	[P value]	Coeff.	odds ratios	[P value]	Coeff.	odds ratios	[P value]	Coeff.	odds ratios	[P value]	Coeff.	odds ratios	[P value]
Sex	-0.68	0.5	0.07	-0.48	0.62	0.15	-0.43	0.64	0.16	-0.29	0.75	0.4	-0.48	0.61	0.13	0.007	1	0.98
log of age	0.24	1.28	0.61	0.56	1.75	0.21	0.21	1.22	0.64	-0.42	0.65	0.35	1.03	2.8	**0.02**	0.51	1.66	0.25
Marital status	0.68	1.96	**0.03**	0.42	1.52	0.14	0.17	1.19	0.53	0.8	2.23	**0.01**	-0.06	0.93	0.79	0.15	1.16	0.58
Education level	-0.29	0.75	0.39	-0.22	0.8	0.45	-0.86	0.42	**0.01**	-0.54	0.57	0.07	-0.16	0.85	0.59	-0.51	0.59	0.09
Experience of illness in last 2 wks	-0.1	0.9	0.78	0.25	1.28	0.43	0.32	1.38	0.32	-0.62	0.53	0.06	0.12	1.12	0.7	-0.04	0.95	0.89
Currently ill	-3.04	0.05	**0.02**	-1.42	0.24	**0.03**	-0.62	0.54	0.29	-0.54	0.58	0.38	-2.24	0.1	**0.01**	-0.83	0.43	0.21
Ever suffered bilharzia	1.02	2.78	0.12	0.78	2.19	0.17	1.22	3.37	**0.03**	0.61	1.85	0.29	0.99	2.71	0.09	0.34	1.41	0.57
Experience with bilharzia	-1.67	0.18	**0.05**	-1.27	0.28	0.06	-1.38	0.25	**0.05**	-1.89	0.15	**0.01**	0.44	1.55	0.500	-0.36	0.69	0.60
Log of Infection intensity	-0.06	0.94	0.74	-0.16	0.85	0.32	-0.18	0.83	0.26	0.01	1.01	0.93	-0.03	0.96	0.84	-0.21	0.81	0.19
Other parasites	0.02	1.02	0.85	0.02	1.01	0.85	-0.09	0.91	0.34	0.26	1.29	**0.02**	0.02	1.01	0.87	-0.08	0.92	0.37
Clinician's diagnosis during hospital visit	0.27	1.31	0.38	-0.26	0.76	0.33	0.36	1.43	0.19	0.29	1.34	0.29	-0.03	0.97	0.922	-0.26	0.77	0.35
Patient complaint during hospital visit	-0.13	0.87	0.45	0.01	1.01	0.94	-0.2	0.81	0.18	-0.04	0.96	0.79	-0.01	0.98	0.93	-0.13	0.87	0.38
severity of abdominal pain and discomfort	0.38	1.46	0.1	0.4	1.49	**0.05**	0.23	1.26	0.26	-0.08	0.92	0.68	-0.03	0.96	0.87	0.11	1.11	0.6
Severity of diarrhoea	-1.01	0.36	**0.01**	-0.68	0.51	**0.03**	-1.13	0.32	**0.001**	-0.88	0.41	**0.01**	-0.46	0.63	0.14	-0.56	0.57	0.08
Severity of watery diarrhoea	-0.21	0.81	0.34	0.23	1.25	0.26	-0.01	0.99	0.97	0.08	1.08	0.69	-0.02	0.97	0.89	-0.36	0.69	0.07
Severity of bloody diarrhoea	-0.12	0.89	0.56	0.06	1.06	0.73	-0.44	0.64	**0.02**	0.14	1.15	0.42	-0.12	0.88	0.49	0.05	1.05	0.77
Severity of tiredness	0.07	1.07	0.73	0.36	1.44	**0.05**	0.33	1.38	0.06	0.25	1.29	0.19	0.15	1.15	0.42	-0.49	0.61	**0.02**
Severity of nausea	-0.11	0.89	0.55	-0.11	0.89	0.49	0.13	1.14	0.45	-0.04	0.96	0.82	-0.001	0.99	0.99	0.16	1.17	0.34

Severity of loss of appetite	*0.28*	*1.32*	*0.17*	*-0.03*	*0.96*	*0.87*	*0.32*	*1.37*	*0.08*	*0.07*	*1.07*	*0.69*	*0.09*	*1.1*	*0.58*	*0.02*	*1.02*	*0.91*
Severity of itching skin rash	0.09	1.1	0.44	0.02	1.02	0.83	0.27	1.31	**0.02**	0.05	1.05	0.61	-0.05	0.94	0.63	0.06	1.05	0.62
Severity of fever	0.01	1.01	0.93	0.03	1.03	0.84	0.28	1.32	0.08	-0.03	0.96	0.81	-0.09	0.91	0.53	0.05	1.05	0.73
Severity of dizziness	0.07	1.07	0.68	0.09	1.1	0.54	0.03	1.03	0.84	0.39	1.49	**0.02**	0.04	1.04	0.78	-0.21	0.81	0.2
Aggregate symptom frequency index	0.24	1.27	0.39	-0.38	0.68	0.11	-1	0.36	**0.01**	-0.23	0.79	0.37	0.39	1.47	0.11	0.99	2.71	**0.00**
Aggregate symptom intensity index	-0.62	0.54	0.18	-0.87	0.42	**0.03**	n.a	n.a	n.a	-0.33	0.71	0.42	-0.19	0.83	0.63	-0.29	0.74	0.47
Aggregate severity index	n.a	n.a	n.a	n.a	n.a	n.a	-1.18	0.3	**0.01**	n.a	n.a	n.a	n.a	n.a	n.a	n.a	n.a	n.a
cut off 1 (_cut1)	-5.22			-5.13			-9.14			-7.37			0.61			-3.35		
cut off 2 (_cut2)	-2.65			-3.73			-7.34			-4.68			1.67			-2.01		
cut off 3 (_cut3)				-1.68			-5.83			-2.78			3.07			0.22		
pr (xb+u<_cut1)	0.2125			0.1			0.05			0.0125			0.1875			0.1125		
pr (_cut 1< xb+u<_cut2)	0.6000			0.2875			0.3125			0.3125			0.3000			0.2875		
pr (_cut2 xb+u)	0.1875			0.5			0.4125			0.5			0.3625			0.5		
pr (_cut3xb+u) for HRQL domains)				0.1125			0.225			0.175			0.15			0.1		
Sample size	80			80			80			80			80			80		
Log likelihood	-51.9			-73.63			-76.07			-65.67			-88.28			-71.74		
LR chi^2	48.12			41.7			42.12			39.83			35.82			45.48		
chi^2	0.0020			0.0139			0.0125			0.02			0.06			0.005		
Pseudo R^2	0.3168			0.2207			0.2168			0.2327			0.1687			0.2407		
Ramsey RESET (p>F)	0.3687			0 942			0.762			0.6943			0.6943			0.1887		
Cook-Weisberg (P>chi^2)	0.1005			0.9205			0.6217			0.1418			0.8594			0.8785		
RESET test (P>chi^2)	0.5246			0.6404			0.6394			0.1888			0.4030			0.0689		

6.5.4.3.1 HRQL and socio-demographic characteristics

Except for feelings of worry and anxiety, the negative coefficients on the other HRQL indicators suggest that females were more likely to be better off on HRQL indicators. However, all the coefficients on gender were not significant. For all models there was a higher chance of older respondents reporting worse off health status indicators. The coefficient on social participation was statistically significant at $p<0.05$, with a high odds ratio of 2.8. There was statistically significant evidence that single respondents were more likely to report worse off on health status index and feeling of strength and energy. However, though statistically insignificant, the single were also more likely to be worse off on other HRQL domains except social participation, where they are likely to be in better off categories. All coefficients on education level were negative indicating that those with lower levels of education were more likely to report adverse effects on these domains. The coefficient on performance of daily duties was statistically significant at $p<0.05$.

Those who experienced illness two weeks prior to measurement of disease states were more likely to be worse off in health status, feeling of energy and strength and worry and anxiety, but more likely to be better off in mobility, performance of daily duties and social participation. None of the coefficients were significant. However, for all the models those ill during the interview were more likely to be worse off in all health status indicators. Coefficients on health status index, mobility and social participation were significant at $p<0.05$. In relation to experience with Schistosomiasis Mansoni, there was very strong evidence ($p<0.05$) for a higher probability that those who had had experience with the illness were worse off in terms of health status index, performance of daily duties and feeling of energy and strength. Other coefficients were negative except for social participation and were not statistically significant.

6.5.4.3.2 HRQL and infection intensity

Contrary to expectation, coefficients on all health status indicators except feeling of energy and strength indicate that those with a higher log of infection intensity were more likely to report better off on HRQL indicators. None of the coefficients were statistically significant. While presence of other parasites was related to a higher probability of being in a worse off health status, mobility, feelings of energy and strength and social participation, only for feeling of energy and strength was this relationship statistically significant.

6.5.4.3.3 HRQL and symptom severity

Table 6.4 summarises the evidence on symptoms and aggregate indices in terms of symptoms and the signs of their coefficients for the health status indicators. Although evidence on all symptoms except diarrhoea, aggregate index on intensity and severity was mixed, the majority of coefficients were positive suggesting that those experiencing severe symptoms had a higher probability of having worse off health status indicators. This evidence was statistically significant for mobility with respect to abdominal pain and discomfort and tiredness, and for performance of daily duties, energy and strength and worry and anxiety with respect to itching skin rash, dizziness and aggregate frequency index respectively. Contrary to expectations, evidence suggested that those with severe diarrhoea, those with higher aggregate intensity and severity index had more likelihood of reporting better health status indicators, with coefficients on health status index, mobility, performance of daily duties and energy and strength being significant at $p<0.05$.

In terms of the hypothesized relationships, there was mixed evidence regarding construct validity of symptoms and HRQL indicators. In most cases there was evidence indicating that the more severe the symptoms the worse off HRQL indicators were likely to be though significant in a few cases. Except for worry and anxiety and performance of daily duties where tiredness and aggregate frequency index respectively, had unexpected coefficients, all other cases with significant but unexpected coefficients involved diarrhoea and overall

Table 6.4: Summary of signs on coefficients on symptom severity by health status indicators

Symptoms	Positive coefficients	Negative coefficients
Abdominal pain and discomfort	Health status index Mobility* Performance of daily duties Worry and anxiety	Energy and strength Social participation
Diarrhoea		Health status index* Mobility* Performance of daily duties* Energy and strength* Social participation Worry and anxiety
Watery diarrhoea	Mobility Energy and strength	Health status index Performance of daily duties Social participation Worry and anxiety
Bloody diarrhoea	Mobility Energy and strength Worry and anxiety	Health status index* Performance of daily duties* Social participation
Tiredness	Health status index Mobility* Performance of daily duties Energy and strength Social participation	Worry and anxiety*
Nausea	Performance of daily duties Worry and anxiety	Health status index Mobility Energy and strength Social participation
Loss of appetite	Health status index Performance of daily duties Energy and strength Social participation Worry and anxiety	Mobility
Itching skin rash	Health status index Mobility Performance of daily duties* Energy and strength Worry and anxiety	Social participation
Fever	Health status index Mobility Performance of daily duties Worry and anxiety	Energy and strength Social participation
Dizziness	Health status index Mobility Performance of daily duties Energy and strength* Social participation	Worry and anxiety
Aggregate frequency index	Health status index Social participation Worry and anxiety*	Mobility Performance of daily duties* Energy and strength
Aggregate intensity index		Health status index Mobility* Energy and strength Social participation Worry and anxiety
Aggregate severity index		Performance of daily duties*

* Shows that the symptom was significant in the respective model at $p<0.05$

severity and intensity index. However, the majority of cases adduced evidence supporting construct validity of severity of symptoms and HRQL indicators.

6.6 Reliability (Internal consistency)

The Cronbach alphas suggest that the symptoms and HRQL sub-scales as well as the total scale were internally consistent as all the alphas were above 0.7. In addition, the inter-item correlations were reasonably low indicating absence of redundancy amongst the items. Item-total correlations for the HRQL domains were above 0.63 and exclusion of any one of the items had the effect of reducing the alpha, which indicates that all the six items improved the internal consistency of the HRQL domains sub-scale. However, item-total correlations for the symptom sub-scale ranged from 0.18 to 0.68. Removal of all aspects of diarrhoea, frequency and intensity of watery diarrhoea and vomiting and frequency of itching skin rash from the symptoms sub-scale and the total scale improved the Cronbach's alphas, an indication that these items were not tapping the same construct.

Table 6.5: Reliability: Internal consistency, inter-item and item total correlations for patient sample (n=80)

	Symptoms: frequency, intensity and frequency of disruption (32 items)	HRQL domains (6 items)	Total: symptoms and HRQL domains (38 items)
Cronbach's alpha	0.85	0.80	0.87
Average inter-item correlation	0.15	0.40	0.15
Item total correlations: range	0.18 – 0.68	0.63 – 0.80	0.12 – 0.73

6.7 DISCUSSION

This discussion addresses the question of whether the questionnaire developed and used amongst the Kikuyu in Kenya to assess the impact of Schistosomiasis Mansoni on HRQL is valid. The discussion starts by examining whether the questionnaire was able to differentiate between patients whose infection status was known and the community members whose infection status was unknown, regarding reports of symptoms and HRQL domains. The discussion proceeds to examine the evidence for construct validity from both correlation and hypothesis testing in order to point out which items of the questionnaire could be considered valid or not and why. The discussion also addresses the question of whether the questionnaire can be generalised to other settings.

The questionnaire differentiated between patients and community members on a number of aspects, where differences between groups were significant. These included; age, marital status, occupation, reports of illness and health problems, knowledge and experience of bilharzia. Also more patients than community members reported having each of the symptom with the most common symptoms being abdominal pain and discomfort, tiredness, fever, nausea, loss of appetite and dizziness that were reported by between 66-91% of patients. Evidence suggests that patients experienced more adverse effects from the symptoms, for longer periods and with higher intensity than community members. They also experienced more frequent disruption of their usual work, household chores and schoolwork due to symptom experience. In terms of HRQL domains, although both groups concurred that all the six domains would be affected due to illness with bilharzia, community members reported being affected less often compared to patients. 81% of the patients had moderate to heavy infection intensity.

That the questionnaire was able to differentiate between patients and community members on the above aspects could be used to argue that the questionnaire assessed the impact of schistosomiasis and therefore was a reasonably valid tool for that purpose. The questionnaire showed that known patients had more experience of symptoms, which

bothered them more and had their HRQL domains affected in various ways. Although the magnitude of reports of symptoms and effects on HRQL domains was more amongst patients, there was also a good deal of community members experiencing them and who noted that the symptoms bothered them and affected their HRQL domains. Hence, this evidence supports the validity of the questionnaire.

Evidence from correlation analysis and hypotheses testing also provided substantial evidence of construct validity of the questionnaire that consisted of symptoms, HRQL domains and assessment of infection intensity. These three measures were expected to correlate in the same direction on the assumption that they were tapping the same latent variable, namely illness with bilharzia and its impact on HRQL. In the majority of cases loss of appetite, tiredness, abdominal pain and discomfort, nausea, dizziness, bloody diarrhoea and fever correlated positively and significantly with most of the HRQL domains, except performance and output of work. Although the majority of the correlations were positive and low, ranging between 0.1 to slightly above 0.5, only those above 0.22 were significant.

Symptoms like abdominal pain and discomfort, fever, dizziness, bloody diarrhoea, watery diarrhoea had low (<0.18) but positive and insignificant correlations with infection intensity. While this was very weak evidence for construct validity it gave a pointer towards the symptoms that had direct relation to infection intensity. It was notable that tiredness did not correlate positively with infection intensity despite being one of the symptoms adducing strongest evidence for validity between symptoms and HRQL domains. There thus appears to be some complex relation between the way infection intensity relates to symptoms and in turn how symptoms relate to HRQL domains. That the evidence for symptoms and infection intensity was weak might be a result of existence of co-morbidities, such that the symptoms were due to other parasites in addition to Schistosomiasis Mansoni.

Therefore it appears from the evidence that the most valid symptoms were loss of appetite, tiredness, abdominal pain and discomfort, nausea, dizziness, bloody diarrhoea and fever,

while all HRQL domains except performance and output of work were valid assessors of how HRQL was affected. The correlation coefficients are however weak to moderate, but statistically significant.

There are a number of reasons that the relatively low correlations should not be interpreted to imply lack of evidence for validity. Brooks (1995) quoting Bombardier and Tugwell (1987) notes; "since there are no gold standards, construct validity implies a comparison of instruments to be tested to other available instruments measuring similar concepts". In such a situation, the correlation between the new and the old measures should be high. In this study, choice and use of symptoms and HRQL domains was new and therefore it could be argued that although they both measured a construct in known patients, there was a possibility that they were tapping other aspects not necessarily related to Schistosomiasis Mansoni. These could result from presence of co-morbidities or other aspects affecting HRQL that were not captured by the questionnaire, hence resulting in low correlations. Another reason for the low correlations was that the construction of the questionnaire followed a clinimetric approach. Fayers and Hand (2002) have shown that when a questionnaire contains causal variables, as did this one, not very high correlations should be expected and low correlations do not necessarily imply lack of validity for the items.

Other reasons for low correlations could be related to lack of input from patients and local community with disease experience from the outset in the conceptualisation of Schistosomiasis Mansoni disease and its impact on HRQL. The fact that construction of the initial questionnaire relied on reviews of literature before the views of patients and health professionals were sought means that previous potential biases in the literature were implicitly contained in the questionnaire. Perhaps the results would have been different (better) had the symptoms and HRQL domains included in the questionnaire been obtained directly from the local people including patients. Also, the patients and health professionals' sample whose views guided the construction of the questionnaire was relatively small and this might have influenced the input into the questionnaire. Future studies should use large samples together with focus groups to get a broader view of how

the local people perceive and consider important concerning Schistosomiasis Mansoni disease and its impact on HRQL.

Contrary to expectation, both correlation analysis, hypotheses testing and regression analysis did not provide any evidence for construct validity between infection intensity and HRQL domains. This implies that infection intensity had no direct relation to how patients' HRQL was affected and therefore it would be unjustifiable to use infection intensity as an indicator of the impact of Schistosomiasis Mansoni HRQL domains. Because the patients were known to have schistosomiasis, one interpretation of this finding is that use of infection intensity as a measure of outcome is partial and narrow. It does not account for how the disease impacts on the patient's HRQL due to symptom experience and bother.

To what extent is this questionnaire generalizable to other settings? This questionnaire has reasonable construct validity in terms of use of symptoms and HRQL domains to assess the impact of Schistosomiasis mansoni. There is also weak indication that most of the symptoms are related to infection intensity as expected, an issue worth further exploration. The fact that symptom frequency was categorised as number of days presents a relatively standard way of assessing symptoms and this can be generalised to other settings. Assessment of intensity of symptoms like loss of appetite was influenced by local ways of expressing how bad the symptom was. Also terms expressing intensity such as 'mild, moderate, severe and very severe' and 'a little, somewhat, very and extremely' may not exist in some local languages or may have no equivalent terms thereby making use of these terms elsewhere require attention. The same is true for terms used to assess how often HRQL were affected, i.e. 'a little, some and most of the time'. It is possible that these terms might mean different things in different cultural settings and languages. Equivalence of such terms would therefore need to be investigated before the questionnaire can be used elsewhere, and for its results to be comparable in different settings. Salomon et al. (2002) have shown that response categories are interpreted differently in different settings. This might influence comparability of measures of the disease between settings and subsequent values attached to disease states constructed from such a questionnaire.

It appeared that symptoms like bloody diarrhoea, watery diarrhoea and diarrhoea were not as prominent as suggested in the literature (Gryseels, 1992). It was not immediately clear why this was the case. Possibilities are that for some reason, patients did not report their diarrhoeas or the literature has been biased in using bloody diarrhoea as an indicator for presence of Schistosomiasis Mansoni. On the other hand, it is also possible that the symptoms reported were not only due to Schistosomiasis Mansoni due to possibilities of presence of co-morbidities, although we were unable to collect data on co-morbidities. Whatever, the reason, it appears that symptom reporting might differ between settings. Such differences restrict the generalizability of results from use of the instrument. Another related issue concerns how different symptoms would be rated in different settings and the extent to which people are willing to publicly report experiencing them. While the questionnaire developed in this thesis offers a template that would be generalizable to other settings, the issues raised would require careful consideration.

One purpose of applying the measurement questionnaire was to create schistosomiasis disease states. Since the symptoms and HRQL domains demonstrated construct validity, it is argued that disease states can be constructed using descriptors both in terms of symptoms and HRQL domains, in terms frequency, intensity and or severity of symptoms and how often HRQL domains are affected. Development of schistosomiasis disease states is described in the next chapter together with developing approaches for valuation of disease states in Kenya.

CHAPTER 7

METHODS FOR DEVELOPING AN APPROACH TO VALUE DISEASE STATES IN KENYA

7.0 INTRODUCTION

Chapter three highlighted the limited use and understanding of performance of both health-related quality of life measurement and valuation instruments in the context of developing countries and amongst parasitic diseases. The previous two chapters were devoted to development and use of a schistosomiasis HRQL measure, which would facilitate construction of disease states for valuation. This chapter, the first of two devoted to valuation of schistosomiasis disease states, deals primarily with pre-testing and choice of valuation techniques for use in Kenya as well as a description of the process of choosing which disease states to value. It aims to present methods for the choice disease specific states for valuation and the development of valuation approaches used in the empirical study reported in the next chapter.

Three valuation techniques, the visual analog scale (VAS), the time trade off (TTO) and the standard gamble (SG) were pre-tested to assess content validity. Content validity was ascertained in terms of: the existence of key terms and concepts; ease of use and understanding; appropriateness; and practicality in a Kenyan rural community. To test the content validity of these instruments, disease states were constructed from the questionnaire used amongst Schistosomiasis Mansoni patients developed in chapter 5.

This chapter begins with a presentation of the approach followed in the construction and choice of disease states used. Section 7.2 presents the approaches to pre-test and justify the choice of valuation techniques used in the valuation study. Results of the pre-test are presented in section 7.3, followed by a discussion in section 7.4. This section highlights the implications of the findings and sets the basis on which valuation techniques and disease states for use in the empirical study are chosen. Section 7.5 presents the conclusion.

7.1 CONSTRUCTION AND CHOICE OF DISEASE STATES FOR VALUATION

Use of a schistosomiasis HRQL questionnaire, developed specifically for this study, and consequent testing of its validity amongst Schistosomiasis Mansoni patients, concluded that disease states could be constructed using descriptors in terms of both symptoms and HRQL domains. Tiredness, loss of appetite, nausea, abdominal pain and discomfort, fever and dizziness consistently provided construct validity while bloody diarrhoea and itching skin rash did so only in some cases. In terms of HRQL domains, performance of daily duties, feeling of energy and strength and social participation adduced most evidence of construct validity while mobility and feelings of worry and anxiety did so in a few cases. Therefore, in constructing disease state scenarios, symptoms and HRQL domains were combined into six dimensions of: mobility, performance of daily duties, output and performance of work, feeling of energy and strength, social participation and worry and anxiety.

Construction of symptom severity indices during fieldwork adopted a multiplicative as opposed to additive approach and correlation coefficients used were Pearson's as opposed to Spearman's and Cramer's V used for the final analysis of use of the schistosomiasis disease states measurement questionnaire. Disease states constructed during fieldwork contained fewer symptoms and included all six HRQL domains as opposed to those resulting from final analysis[57] as presented in appendix 7.1. The main differences are: inclusion of bloody diarrhoea; exclusion of fever, nausea, abdominal pain and discomfort and dizziness; inclusion of the HRQL domain performance of output and work in the disease states constructed during fieldwork compared to those contained in appendix 7.1. Considering that the exercise was more methodological and focused on valuation instruments rather than empirical focusing on measuring disease states, it can be argued that these differences are unlikely to influence methodological concerns greatly.

[57] At this point attention is drawn to some differences in the construction of symptom severity indices and therefore the analysis correlation coefficients during and after fieldwork, which led to some minor differences in the disease states constructed and used in the empirical study and those subsequently resulting from the final analysis.

To keep an interview to a manageable time, the number of states to be valued was set at 8. The first step towards choice of these states was selection of 6 states from patients' reports of how Schistosomiasis Mansoni affected the six selected HRQL domains on three levels. To ensure an even spread and representativeness, the 17th, 33rd, 50th, 67th, 83rd, and 100th percentiles of the frequency distribution of six HRQL domains reported by the patient sample were selected. The states 'perfect health' (denoting no symptoms and no effects on HRQL domains) and 'dead' were included to act as endpoints and create 8 states. The six were as follows:

Table 7.1: Six selected states as represented by HRQL domains on three levels

Percentile	Corresponding levels of the six HRQL domains	Alphabetically labelled as:
17	111111	A
33	122232	B
50	213102	C
67	222222	D
83	232222	E
100	333333	F

To draw statements describing the above states in terms of symptoms and HRQL domains the 25th, 50th and 75th quartiles were computed for domains and symptoms that were found to have statistically significant associations. This facilitated drawing out three statements describing how and by which symptoms each domain was affected on three levels namely; a little, some and most of the time (see Appendix 7.2). For example, through construct validation, there were significant correlations between mobility and tiredness and fever. From appendix 7.2, the 25th, 50th and 75th quartiles corresponding to mobility, tiredness and fever were as specified below.

Table 7.2: Example of construction of descriptors for disease states: Mobility HRQL domain

Percentiles	Domain and levels 1, 2, 3	Symptoms			
	Mobility	tiredness		fever	
		Frequency	Intensity	Frequency	Intensity
25	1=(a little of the time)	2=(1-3 days)	2= (somewhat)	2=(1-3 days)	2=(moderate)
50	2=(some of the time)	3=(4-6 days)	3=(very)	3=(4-6 days)	3=(severe)
75	2=(some of the time)	4=(7-10 days)	3=(very)	3=(4-6 days)	3=(severe)

From this table, statements describing how mobility would be affected on three levels by symptoms were drawn out as follows:

MOBILITY

1. You feel somewhat tired for 1-3 days; have moderate fever for 1-3 days and this affects your mobility a little of the time.

2. You feel very tired for 4-6 days; have severe fever for 4-6 days and this affects your mobility some of the time.

3. You feel very tired for 7-10 days; have severe fever for 4-6 days; and this affects your mobility some of the time.

This procedure was replicated for each of the six HRQL domains and then used to construct the six disease states each described on appropriate level for each domain (see appendix 7.1). It was noted that after specifying each domain in relation to symptoms significantly affecting it at different levels and then combining this into a scenario, there was a lot of repetition of symptoms and at times different levels of symptom severity would appear in one scenario. To overcome this problem all symptoms appearing in a scenario were summarized, taking into consideration the levels at which various health domains were affected. The scenario was then represented as a summary of symptoms and the health

domains (see example below). The full description of the six selected states was as presented in appendix 7.4.

Example of a Schistosomiasis Mansoni disease state scenario

State 111111	A
♦ You feel somewhat tired for 1-3 days ♦ You can only eat about half to a quarter the amount of food that you normally eat for 1-3 days ♦ Your mobility, performance of your daily duties, performance of output and work, feeling of strength and energy in body, social participation are affected a little of the time and you feel worry and anxiety a little of the time.	

These disease states were colour-coded green to denote that they were states with some problems. States 'perfect health' and 'dead' were colour-coded pink and blue to distinguish them from the states being valued.

7.2 METHODS FOR CONTENT VALIDATION OF VALUATION TECHNIQUES

This section presents the methods and procedures followed in pre-testing three valuation techniques namely, VAS, TTO and SG. It starts with a presentation of sample selection, followed by description of construction of the props used in the exercise. The pre-test procedure and issues addressed in content validating each instrument are finally presented.

7.2.1 Sample and sample selection

A convenient sample of 28 respondents was selected from Mwea. The sample was selected in a way that would have the views of the old (above 35 years) and young, males and females represented. Seven respondents from each group were selected.

7.2.2 Instruments and props

Considering the novelty of the use of health states valuation techniques in this community, it was deemed necessary to use visual aids and props in the valuation exercise. These included colour-coded laminated health state cards. The props included a VAS board, TTO board and a SG chance board measuring approximately 41 by 56 centimeters each (see appendix 7.3). Construction of these props followed Furlong et al. 1990's guide to design and development of health-state utility instrumentation.

7.2.3 Pre-test procedure

Training of interviewers: The principal researcher (PR) did very careful and intensive training of the interviewers. This involved explanation of the concept of health state valuation and how each instrument is used to perform the valuation task. Using the props, the PR demonstrated how each instrument is used by going through a trial interview, and allowing the interviewers to ask questions and seek clarification on issues that were not clear. This continued throughout the training sessions. The interviewers then practiced among themselves with the supervision and guidance of PR until the interview script, use of props, procedure and purpose were completely mastered by heart by all. Each interviewer practiced gaining an interview until the introduction was deemed to be natural. Then each interviewer practiced each method while the rest observed and gave feedback. This was repeated until the method was completely understood. The questionnaire and the valuation interview script were studied carefully before a collective translation of the instruments into the local language, Kikuyu.

Translation procedure: Owing to financial limitations, the standard recommended procedure of forward and back-translations involving teams in both original and target cultures were not followed. Translation of the instruments was done by the research team (comprising two sociologists and one linguist, all University of Nairobi graduates) recruited from the study area together with the principal researcher. The advantage of translators from the study area was that they spoke the local dialect of the Kikuyu language, that was

used in interviews. All items in the questionnaire were translated one at a time. The translation started with each person doing an independent translation. The independent translations were then compared and debated upon by the team to determine and build a consensus on which translation conveyed the exact meaning of the item. Examples of differences in translations mainly regarded sentence syntax (where a translator translated the item word for word) which conveyed slight differences in meaning and use of different words that have similar meanings in Kikuyu. For instance, several local words for worry and anxiety i.e. *kimako, gitangiko, kieha, gwitigira, gutuikira* can be used depending on the context and locality. During this process, input from the local research assistants was considered vital in selecting local terms that would convey the intended meaning. The agreed version of the item was then adopted. This process was deemed necessary for ensuring uniformity during interviews and it also contributed to better understanding and grasping of the instrument by the research assistants.

Pre-testing Format: Early on, it had been established that pre-testing the three methods on each respondent would take nearly three hours and was extremely exhausting for the respondent. It was therefore decided that the methods would be pre-tested on the 28 respondents as outlined in table 7.3. This meant that each valuation technique was pre-tested once on 16, twice on 12, and thrice on four respondents spread equally across age categories and gender.

Table 7.3: Distribution of respondent category by valuation technique pre-tested

Valuation technique	Respondent category				Total
	Young male	Old male	Young female	Old female	
VAS	1	1	1	1	4
TTO	1	1	1	1	4
SG	1	1	1	1	4
VAS, TTO	1	1	1	1	4
VAS, SG	1	1	1	1	4
TTO, SG	1	1	1	1	4
VAS, TTO, SG	1	1	1	1	4
Total	7	7	7	7	28

Interview Procedure: Pre-testing the valuation techniques took place in two parts; taking the respondent through one valuation task followed by a debriefing about their use of the method. To save on interview time only one state (labeled C and typed on a green laminated card) was valued while using perfect health (pink) and dead (blue) as endpoints for each of the techniques.

7.2.4 Issues addressed in content validation

Evidence on content validity of the instruments is often lacking in the original and target cultures. At best reliability and construct validity are assessed and this is used as a basis to facilitate comparison of values across cultures. It is argued that before examining reliability and construct validity, an examination of content validity should be part of the evidence basis for judging and interpreting cross-cultural comparisons of values (McDowell and Newell, 1996; Herdman et al. 1997 and 1998; Carmines and Zeller, 1979). An in-depth structured interview was developed to examine the VAS, TTO and SG in four main ways; whether terms and concepts embodied in each technique existed in Kikuyu; ease of understanding and use of the techniques; appropriateness of the techniques to the community, and practicality. Each of these are described in detail below.

7.2.4.1 Terms and concepts

Terms and concepts embodied in each instrument were singled out. The PR examined and singled out the concepts embodied in each instrument as well as the key terms encountered in the process of valuation, by thinking through the process of obtaining a value/utility using each technique and the questions posed to the rater. For instance in using the VAS, the respondent is asked to visualize or imagine themselves a state scenario and then rank it in relation to others as a way of expressing their preference in terms of a number. This process was repeated for each instrument to isolate the concepts, key terms and tasks the respondent undertook in searching for a value using the technique. During the debriefing of the use of each method, the interviewer pointed out to the respondent the concepts and key

terms that they encountered. The respondents were then asked to talk about each concept or key word giving instances and examples of existence of these terms and concepts. Assessment of the terms and concepts was general rather than restricted to health issues and responses were expected to be broad covering different aspects of day to day experiences of the respondent. It was assumed that if a concept or key term was non-existent, then there would be no examples of instances relating to the use of he concept or the key word. It is worth noting that some concepts are common to all the instruments although they were not assessed in relation to all, to avoid repetition of the same question to the respondent.

The terms and concepts selected for each instrument were as follows. For the VAS technique the terms and concepts included: ranking a health state; expressing preference in terms of a number; imagining oneself being in a scenario; and the use of a VAS thermometer to rank health states. For the TTO technique the terms and concepts included: giving up time; making choices; concept of future and how expressed; perfect health; making comparisons between lives and choosing one; concept of time; concept of life; idea of giving up time in bad health to stay in good health; talking about death. With respect to the SG, the terms and concepts included: risk taking; concept of chance or probability; gambling; uncertainty and how it is expressed; idea of risking immediate death; idea of gambling with two lives and taking a chance of dying.

7.2.4.2 Ease of understanding and use of method

This was assessed in two ways. After explaining the valuation task for each method, the interviewer assessed whether the respondent understood the task and also recorded the number of times the task was explained before the respondent understood. Secondly, the respondent was asked to talk about any difficulties they experienced in going through the valuation task and to rate on a 5-point scale (ranging from very easy to understand through very difficult to understand), how they found each valuation procedure. For respondents evaluating two or more techniques, they were asked to state their preferred technique giving

reasons for their preference as well as their recommended technique for use in that community.

7.2.4.3 Appropriateness

Asking the respondents' opinions regarding the valuation techniques in terms of five areas assessed appropriateness. These were whether there were any questions that would: annoy; cause embarrassment; convey the wrong meaning; sound disrespectful; cause problems and make someone upset or feel offended. For each aspect the interviewer probed for details regarding which aspects of the techniques might be inappropriate.

7.2.4.4 Practicality

Practicality was assessed through a variety of ways. First we tested the respondents' capability of differentiating between two scenarios. Two disease states (A and D) were described to the respondent and asked which of the two was better off and why. They were also asked to point out the differences between the two states and whether they had any difficulties imagining themselves in the two states. The respondents were also asked to indicate on a 4-point scale (from very hard to very little), the amount of thinking required for each valuation task. In addition, they were asked to talk about the things they worried about, strongest feelings, thoughts or emotions they experienced during each task. Finally, they were asked to state all the different types of people who should not be asked to undertake a valuation task and why. Appendix 7.7 presents a complete interview schedule for pre-testing the three valuation instruments.

7.2.5 Methods of analysis

Analysis was limited to frequencies and descriptive statistics, comparing across techniques for each of the four areas assessed. Regarding existence of terms and concepts, opinions will be reported as text making references to frequency of opinion amongst respondents

where applicable. This will focus on presenting arguments for and against recommending or not recommending use of the technique in this setting.

7.3 RESULTS

7.3.1 Respondents characteristics

The mean age of respondents was 40 years (SD; 15.6 years) and ranged between 21- 74 years. 75% were married and had children as opposed to 21.4% single and 3.6% widowed who had none. Those reporting educational levels of primary and above were 82.9% of whom 61.4% had post secondary education. The majority (71.4%) relied on farming to earn a living although nearly 50% were also engaged in other activities such as teaching, civil service jobs and business.

7.3.2 Existence of terms and concepts found in valuation techniques

To avoid repetition, terms and concepts common to all instruments are presented together in section 7.3.2.4. Sections 7.3.2.1 to 7.3.2.3 presents terms and concepts in VAS, TTO and SG respectively. As noted earlier, the following reports refer broadly to the concepts in general.

7.3.2.1 Visual analogue scale

Ranking: All respondents gave instances and examples of situations where people in the community use ranking. These examples related to comparing and ranking farming related activities (56.2%), comparing and ranking aspects in life (31.2%) and ranking and ordering daily activities (12.5%). For example, in farming people *"..... rank activities in order, say preparing the land, planting, cultivating, harvesting, and so on"*. Also, in farming "*people rank activities they carry out depending on which earns them quick money. For example french beans come first, then tomatoes, then rice and other food crops.*" Examples

of ranking aspects in life included "*ranking household chores in order, say preparing food and children for school, and then other daily activities*" as well as other aspects such as "*education, farming etc depending on whether one has resources to do these things.*"

Expressing preference: With respect to expression of preferences the majority of respondents gave examples of preference relating to farming (43.8%) and lifestyle (43.8%). For example, farmers express preference for certain crops by planting them while in lifestyle those who love life migrate to big towns where there are many attractions. It was also apparent from the examples that people prefer that which they perceive as being beneficial or deserving higher priority.

7.3.2.2 Time trade off

Choice: The concept of choosing was expressed in terms of choices regarding lifestyle (62.5%), choices related to farming activities (31.25%) and choices related to occupations (25%). Choosing to be educated, lead a morally upright life, work hard rather than live in poverty were some lifestyle choice examples, while in farming choices involved which crops to plant and peoples' choices of rice marketing. It appeared that in making these choices, people considered the ones with more gains and benefits and also drew from their past experiences.

Giving up: To assess the concept of giving up, respondents were asked to give examples and instances where people in the community give up one thing for another and how they go about it. Several examples were given including young people giving up on education to become farmers or do other less attractive activities (37.5%), giving up growing one crop in favour of another (37.5%) and giving up married life for single life (25%). Others included giving up on farming to migrate to towns (25%), giving up on the NIB (National Irrigation Board) in favour of the Society for rice marketing (18.5%) and giving up on life by engaging in sexually immoral activities that can kill (12.5%). The concept of giving up seemed to be common in the community and an exact local word *gwicama*, existed to describe the concept.

Future: The concept of future was found to exist and was expressed in terms of time to come (from now onwards) (81.25%), in terms of activities they are engaged in and would like to undertake (25%) as well as in terms of when they grow old (18.75%). Conceptions of the future were associated with making plans or undertaking activities that would take care of the time to come such as educating children, building a home. In reference to the future, people use the term *"thinking of tomorrow"*. There is also an aspect of uncertainty when people talk of the future, as evidenced by constant reference of '*you can die at any time', 'when you have gone', 'time belongs to God' 'anything can happen in between', 'you can go anytime*'. All these were references to the fact that though they think and plan about the future, there is no telling if your plans will materialize. When asked how far into the future they think, the responses were varied. Some expressed in terms of months (31.25%), years (25%) while others referred to the rest of their lives (18.75), with some giving it a time limit in years. However, the majority (50%) of respondents mentioned that considerations of the future depended on planned activities and that since the majority of plans revolve around rice farming, most people would identify with thinking of future in terms of one year.

Time: To assess the concept of time and how people relate to time, respondents were asked to say how people in the community talk and think about time. All the respondents mentioned that people think of time in terms of activities planned for the day while 50% referred to time in terms of durations of portions of time in a day. The activities range from daily chores to farming activities which take up to a year.

Life: The concept of life is constantly encountered in TTO as respondents are asked to choose one life or the other. Respondents were asked to say how people talk and think about life. Responses included how their lives are regarding the activities they are involved in while alive (56.25%), in terms of level of comfort while alive (good life versus life of poverty) (18.75%) and how well one is settled in terms of property ownership and personal outlook. Examples of how people talk and think about life included, *"……..in terms of the work they do, say employment or farming…….. Life is from when I was born until when I die and the environment I live in and things I do while alive"*, young male and female. In

terms of comfort while alive two young males and a female remarked, *"...when you have health and money, that is good life. Nothing is bothering you. You know poverty is like a disease? It depends on how you live. Without problems of money, your body is not sick and if sick you have money to go to hospital, ...a good house, your children go to school and you don't struggle to pay school fees and you can give them what they like."* An old man said *"you can look at someone's physical appearance and tell how they have lived"*. The young therefore thought more in terms of the future while the old looked back into the past in thinking about life.

7.3.2.3 Standard gamble

Risk: Several examples were given where people take risks in the community. Most risks relate to investments in development projects such as building, business and buying property (56.25%); risks related to farming decisions (31.25%); behavioural risks such as immorality, stealing, engaging in witchcraft (25%); health related risks associated with sanitation and personal hygiene (18.75%) and risks related to investing in children's education (18.75).

Probability or chance: Respondents were asked to tell of situations where community members would be involved in situations where they have to take a chance or deal with probability of an occurrence. The majority (62.5%) mentioned situations in farming. In farming when they plant there is a probability that they will not harvest depending on seasons and crop diseases and that the prices can be such that they make a loss or gain. The respondents also mentioned that life as a whole is a gamble (37.5%). People do things to succeed and not fail, but only "*God knows the outcome*". Other situations mentioned include balloting in women's groups and land buying companies where they ballot for who gets the money and plot numbers. In this case it means that the group members do not know beforehand who would get the money or which plot and they all have same chance success and failure. These examples embody both the concepts of objective and subjective probabilities.

Gambling: Examples of instances where people are involved in a gamble included farming (68.75%), in life decisions where life itself is seen as a gamble (31.25%) and in women groups' activities (12.5%), where they contribute money and then ballot to decide who receives the money. This was seen as a gamble in two ways. First, because each member of the group has the same chance of receiving the day's contribution and it is not known a priori who that is, and secondly the gamble has another dimension in that it is uncertain to those who have not received whether those who have already received would continue to contribute. In farming, aspects of gambling included being uncertain about the weather and whether one will get a good harvest or not and whether the prices for their crops will be high or low. In life, an example of gambling was such as where parents invest a lot of money educating their children, but they are uncertain of whether the children will get jobs after completing education as well as whether the children will actually complete their education. Besides these, examples of various types of games and activities that both young and old participate in which involve gambling were given. They included *ndia ruru* (a game played using bottle tops by men), *Merry-go-round* for women and *karata* (a game of cards similar to solitaire where you win or lose money). *Karata*, also referred to in Kiswahili as *pata potea* (you gain you lose) is often run by young men (in collusion) where unsuspecting, and often uneducated persons (mostly women) gamble to double their money and end up losing all of it. Although the concept was established to exist, a 29 year old male commented that he would only gamble if the probability of death was 1% while one young and two old females said they would only gamble if the states were so bad. Two young females and an old male failed to provide responses.

Uncertainty: Uncertainty was associated with future events whose outcomes were unknown. The majority (43.75%) provided examples of uncertainty in farming such as weather and future of rice marketing while 37.5% provide examples related to uncertainty in life such as security of land tenure and employment after schooling, and biblically whether there is life after death. Other examples related to uncertainty in the political systems (12.5%) and in health related issues (18.75%) such as doubts about the existence of AIDS. The respondents also provided local phrases used in expressing uncertainty over

something. They included *nganja* (doubts, not knowing how reliable something is, not sure) and *kwigereria* (chancing or probable).

7.3.2.4 Terms and concepts common to VAS, TTO and SG

Imagining: Ability to imagine or visualize oneself in a scenario is vital to health state valuation because valuers often value states they are not in. Respondents were able to relate to the concept of imagining. Imagining was expressed in terms of visualizing oneself in a given situation in time to come (68.8%) as well as developments related to family welfare (18.8%). They used terms such as *"I see myself....."*, *"I visualized myself as....."* and in terms of *"dreams in my mind..."* while others used the term *"I imagined myself..."* itself.

Health: To find out if people could relate to the term health, they were asked to spell out the different ways that people in the community would talk about the term health, expressed as *ugima wa mwiri* in Kikuyu. A variety of descriptors were used the most common being physical appearance (fat, thin, strong, weak, smooth skin, physical maturity of body organs, bodily heat) (62.5%), absence of pain (50%) and without any symptoms or illness (37.5%). Other ways people talk about health that were less mentioned included ability to do daily duties, eat well, without any financial problems and without stress. In addition to the above descriptors, the words *"afya"*, *"helothi"* and a variety of local Kikuyu phrases were used in talking about health.

Perfect health: Although in the VAS the upper endpoint is described as the best imaginable health state, the concept of perfect health is more common in the TTO and the SG. To find out if people could relate to this concept, respondents were asked to say how people talk about perfect health and what they would consider this to be. Their responses are summarized in table 7.4.

Table 7.4: Descriptors of perfect health

Descriptor	% mentioning descriptor (n=16)
Without any symptom or illness	87.50
Without pain	68.75
Ability to eat well and not worried where to get food	50.00
Physical appearance (strong, good body, being mobile, without disability, sexual maturity)	43.75
Without stress, content, psychologically whole, no worries	37.50
Ability to perform daily duties or activities	25.00
Having financial resources to meet your needs	18.75
Spiritual richness and wholeness	6.25

Top amongst the qualities associated with perfect health were absence of symptoms, illness, pain, ability to obtain food and eat well, complete in physical appearance and without worries and stress. Although these qualities were associated with the concept of perfect health and therefore suggest that the concept can be conceptualized in this community, there were five instances where the respondents commented that perfect health does not exist. Some of these comments were: *"... there is nothing like perfect health. That's why I give it 90"*, young man aged 35 years; *"I don't see how somebody can be in perfect health"*, old woman aged 53; *"No one can be 100% healthy. People revolve around 80, 90. You cannot have perfect health. There will always be something to make you stay below 100."*, old man aged 68 years; *"A person in Mwea can never be in perfect health. So I would give it 90. .....You can never be 100."*, young man aged 27 years. *"It is not practical to be in perfect health. You can never remain in perfect health for a whole year, so telling someone that they will be in perfect health is a lie"*, young woman aged 28 years.

Type of comparisons made: The concept of comparison was well understood and respondents gave various instances where comparisons are made. These included comparing different types of lives (75%) such as present versus past and rich versus poor; comparisons of occupations (25%) such as farming and business; and comparisons of framing activities (12.5%) such as different crops and seasons.

7.3.3 Ease of understanding and use of valuation techniques

Table 7.5 summarizes assessment of ease of understanding and use of the three valuation techniques. According to the interviewers' judgement, all respondents understood the valuation task using the three techniques. In all the methods, the same number of respondents had the task explained only once for them to understand. According to respondents' assessment of ease of using the techniques, VAS was deemed easiest, followed by TTO and SG, a pattern reflected in the percentage of respondents regarding the techniques as difficult.

Respondents were asked to state any difficulties they may have had experienced in using the techniques. While the majority experienced no difficulties with any technique, there were more experiencing difficulties with TTO and SG (table 7.5). Difficulties associated with TTO were: comparing time in states and balancing them, questions sounding similar and hence difficulty in understanding what the researcher was after, mentioning death too many times, fearing that what you chose might come true and requiring a lot of thinking to understand the choices. Difficulties associated with SG included: dealing with uncertainty and fear of gambling with death, difficulties in terms of thinking more because it was in probabilities, confusing the gamble because of changing probabilities, and not knowing what would happen in the end. The respondents noted that it may take time to understand, but it is better to try it out on people for them to understand. The only difficulty mentioned with respect to VAS was imagining how a state would be like. They suggested use of names for scenarios to differentiate them more easily.

Table 7.5: Ease of use and understanding of technique (%)

		Visual Analog Scale	Time Trade Off	Standard Gamble
% understanding task (n=16)	Yes	100	100	100
No. of times task explained (n=16)	1	87.50	87.50	87.50
	2	12.50	12.50	6.25
	3	0	0	6.25
Ease of understanding (n=16)	Very easy	6.25	0	0
	Easy	68.75	56.25	50.00
	Fairly easy	12.50	18.75	6.25
	Difficult	12.50	25.00	37.50
	Very difficult	0	0	6.25
% experiencing difficulties during task (n=16)	Did not experience	87.50	56.25	68.75
	Experienced	12.50	43.75	31.25
Ranking of technique (n=12)	1	83.33	41.67	16.67
	2	8.33	50.00	66.67
	3	8.33	8.33	16.67
Technique recommended (n=16)		56.25	50.00	18.75

All the respondents on whom two or three techniques were pre-tested were asked to rank them in terms of 1, 2, and 3, where 1 was the most preferred. The majority (83.33%) ranked VAS first. Although SG was ranked second by more (66.67%) respondents than TTO, more (41.67%) ranked TTO first than SG. Hence, in terms of those ranking TTO and SG first and second, TTO was preferred to SG. In addition, more respondents ranked SG third. Therefore, this ranking reflects a similar pattern to prior assessment. Finally respondents were asked to recommend a technique that should be used in this community. The most recommended was VAS, TTO and SG in that order. Table 7.6 summarises the reasons provided for and against preferring a given valuation technique.

Table 7.6: Reasons for and against preferring valuation technique (n=12)

	Reasons for	Reasons against
VAS	• Simple and easy to understand takes less time (83%) • Has fewer questions and less confusing (17%)	• Difficult to give a state a number just using your mind (8.3%)
TTO	• Easy to understand and make a choice as they could associate with the number of months they would be in a state (42%) • More basic than SG (25%)	• Complex and harder than VAS (25%) • Had many questions and things to think about such as changing months and states (33%)
SG	• Easy to understand (25%) • Complicated due to gambling but better than VAS and TTO (8.3%)	• Confusing due to ping-ponging chance cards therefore difficult to follow (33%) • Questions were trickier than in VAS and TTO (42%) • Difficult to deal with uncertainty and gambling (17%)

In comparison, the majority preferred VAS for its simplicity and ease of understanding compared with those preferring TTO and SG for similar reasons. The only difficulty associated with VAS was assigning a number to a state. Although both TTO and SG were considered more complex and confusing, with more and trickier questions, there were more respondents citing these difficulties for SG than for TTO. Most of the difficulties cited for all the techniques had to do with the process of searching for a value for a given scenario and the mental effort required undertaking the tasks.

7.3.4 Appropriateness of valuation techniques

Table 7.7 shows the proportion of respondents considering use of the VAS, TTO and SG appropriateness according to six attributes used in its assessment. No instrument had aspects that were considered disrespectful. In general, the majority of respondents did not say they considered anything inappropriate about VAS (range: 68.75 -100%), TTO (range: 62.5 – 100%) and SG (75 – 100%). However, a few felt all the techniques had aspects that could annoy, embarrass and convey the wrong meaning. In terms of aspects considered to cause problems, offend or upsetting the VAS came out best and TTO the worst.

Table 7.7: Attributes of appropriateness by valuation techniques (%)

No aspects of an instruments would:	% considering technique appropriate on attribute		
	VAS	TTO	SG
Annoy	68.75	62.50	75.00
Embarrass	81.25	100.0	81.25
Convey wrong meaning	75.00	68.75	75.00
Sound disrespectful	100.00	100.00	100.0
Cause problems	100.00	87.50	93.75
Offend and upset	100.00	93.75	100.00

Appendix 7.5 summarises the aspects of different attributes considered inappropriate for each valuation technique, showing against each aspect the proportion of respondents holding that view. Annoying aspects in all instruments were mostly related to the issue of mentioning death especially to the old and sickly, and people mistakenly thinking that they will be in the states they were asked to value. It was noted that constant mention of death might provoke memories of family members who had died, represent loss of hope in life and hasten one's death, while others may become suspicious of devil worshipping and think you will bring death to them. It was suggested that instead of mentioning death it would be better to say you do not know what will happen after the time in a state. Embarrassing aspects for VAS were related to misunderstanding of the task due to illiteracy or youthfulness, while for SG this related to the interviewer-interviewee age difference with respect to discussing symptoms like bloody diarrhoea[58]. The issue of death was again cited with to respect aspects that could convey wrong meaning with respect to TTO and SG. In general about half of the respondents on whom any instrument was pre-tested expressed the different opinions about inappropriate aspects. In VAS, these were three young males and females and two old males (age range:21-68 years), while for the SG they were both young and old males and females distributed almost equally (age range: 28-74 years). In TTO, those who expressed opinions about inappropriate aspects were mainly young males and females (age range: 27-37 years). It is worth noting that all inappropriate aspects in TTO and in three instances for SG related to the issue of death.

[58] The old people may feel uneasy discussing their diarrhoea with a young person.

7.3.5 Practicality in use of valuation techniques

7.3.5.1 Differentiating between scenarios

Respondents were asked to state, which of two disease states (A and D) was better, and to point out the differences in the two, where a maximum of three differences was expected. All the respondents judged A better than D. Table 7.8 shows the number of times scenarios were described by the number of differences noted. The mean number of differences noted was 2 with only 19% noting all three. The description of scenarios had to be given two times on average, with nine out of 14 people for whom the scenario was described 1-2 times noting 2-3 differences. This implies that the differences were not immediately apparent with one description and therefore could not be noticed at once.

When asked whether they had difficulties imagining and visualizing themselves in the two states, the majority (87.5%) said no. Those with difficulties attributed them to the fact that the descriptor for HRQL domains was a rather long sentence for which shortening was suggested. In this sentence that contained how the six HRQL domains would be affected only performance of daily duties and work and ability to socialize and were noticed by two or more respondents. The rest of the descriptors were noticed only once. In this case the commonly noted differences related to ability to eat, performance of daily duties and symptoms like tiredness.

Table 7.8: Ability to differentiate between two scenarios (% of respondents, n=16)

No. of differences noted	No. of times scenario description repeated			Total
	1	2	3	
1	12.50	18.75	6.25	37.50
2	12.50	25.00	6.25	43.75
3	12.50	6.25	0.00	18.75
Total	37.50	50.00	12.50	100.00

7.3.5.2 Other aspects of practicality

Effort required to use the valuation techniques was assessed in terms of extent of thinking, rated from very hard to very little. The results are presented in table 7.9.

Table 7.9: Extent of thinking required in using technique (%, n=16)

	VAS	TTO	SG
Very hard	6.25	0.00	6.25
Hard	37.50	68.75	37.50
A little	56.25	31.25	56.25
Very little	0.00	0.00	0.00

Both VAS and SG required less effort in that majority thought a little in undertaking the task as opposed to TTO where majority thought hard to do the task. However, about 6% thought very hard for both VAS and SG. None of the methods required very little thinking.

Respondents were asked to state those things they worried about, strongest feelings, thoughts or emotions that they experienced during use of the valuation techniques. Table 7.10 presents the proportion of respondents in whom use of valuation techniques provoked no worries, feelings, thoughts or emotions.

Table 7.10: Proportion with no worries and feelings, thoughts or emotions provoked by use of techniques (%, n=16)

	VAS	TTO	SG
Things respondent worried about	81.25	62.50	75.00
Feelings thoughts or emotions provoked	68.75	75.00	50.00

Respondents tended to be least worried by the VAS, followed by the SG and then the TTO. Views about the VAS and SG swapped order in terms of provoking feelings, thoughts and emotions. Appendix 7.6 presents the aspects that provoked worries, feelings and thoughts. The TTO had the majority (43.75%) of respondents with worries, feelings and thoughts

provoked while SG and VAS had 37.5% and 31.25% respectively[59]. These opinions were from the whole spectrum of respondents. Thinking about death caused worry in all three techniques. Also imagining oneself in death and some severe symptoms caused worry "*because they are bad states*". In SG, due to gambling, thinking that one could land in death caused worry while thinking that one could land in perfect health caused happiness. In TTO having to think about the two lives, and wondering how one could avoid being in the problematic one caused worry. Others were worried about how the problems in the scenarios could be alleviated.

Feelings, thoughts and emotions provoked by VAS related to thinking of dead as 0 and 100 as perfect health. One respondent said that this is unimaginable, noting *"perfect health? .....may be in heaven. Even death we do not know how it is. We live in between but to imagine 0 and 100.....you don't know how they are."* Others thought that it was their health states we were trying to value, but later understood but cautioned that people in the villages may not like to know where their states of health would fall. Others thought how their states could be compared to the ones we presented.

7.3.5.3 People to whom the techniques should not be administered

Table 7.11: Opinions about those who cannot use techniques (%, n=16)

	VAS	TTO	SG
Very old	93.75	81.25	81.25
Young children	75.00	93.75	81.25
Illiterate persons	18.75	12.50	12.50
Very sick persons	31.25	37.50	12.50
Mentally unstable	12.50	31.25	25.00
Others	12.50	0.00	12.50

Table 7.11 shows the groups of people whom the respondents suggested could not use these techniques. They included the very old (over 70 years), young children (6 to 18 years), illiterate persons, very sick and mentally unstable persons. Others included those unwilling

[59] These percentages represent all respondents who expressed worries, feelings and thoughts irrespective of whether they expressed one or two aspects that were worrying or provoking thoughts and feelings. Hence they do not tally with those in appendix 7.5 where responses are not mutually exclusive.

to help, those who have experienced death recently and drunkards among others. The majority of respondents suggested that the very old and young children could not use the three techniques. The very sick persons were mentioned in connection with VAS and TTO while the mentally unstable were mentioned in connection with TTO and SG.

The most common reason that the old, young children, the sick and illiterate persons could not use the various techniques was lack of understanding due to poor memory, age and lack of experience which was cited by 75%, 56.25% and 62.5% respondents with respect to VAS, TTO and SG respectively. A common reason for not asking the old and the very sick was that they might think they would die soon (VAS-18.75%; TTO-25%; SG-12.5%) or feel sicker (VAS-12.5%; TTO-12.5%). It was considered that children might be frightened by mention of death (VAS-18.75%; TTO-12.5%; SG-12.5). The mentally unstable might provide unreliable responses.

7.3.5.4 Opinions about instrument-specific tasks

The rest of assessment of practicality concerns aspects that are both common and specific to instruments. Respondents were asked their opinions about tasks involved in valuation related to each technique. Aspects common to all instruments are presented first followed by those specific to TTO and SG. No tasks were found to be specific to VAS only.

7.3.5.4.1 Tasks common to VAS, TTO and SG

Placing a numerical value on a health state: Asked their opinion about placing a numerical value on a health state, the majority (75%) had no problems and said they considered the problems in a scenario and placed it accordingly. The rest said it was not easy because they had to visualize and it is not something you can see and that one can only rank themselves and not others because one can imagine themselves in a given situation. This view appears to contradict suggestions that it would be inappropriate to ask people to imagine themselves being in these states, expressed regarding appropriateness of techniques, despite it being a minority view (12.5%).

Death: The issue of death appears in the three valuation techniques in different forms. In the VAS respondents imagine or think of death when presented to them and they have to rank it in relation to others states. In the TTO, death is mentioned frequently as the endpoint after the specified duration in the two choices, while in the SG every gamble involves being returned to perfect health or dying immediately. Respondents were asked their opinions about being asked to imagine being dead and risking death.

The majority (VAS-56.25%; TTO-75%) held that it was not bad because everyone knows that they will die one day and dying is a normal event in life. When asked whether asking about death can cause death, these noted that talking about death per se would not cause death in the community. However, they cautioned that people's reactions would depend on how the questions are framed as well as how one talks about death and suggested that the interviewer makes sure that at no time should the respondents think they are the ones in these states. Reactions would also depend on the type of death and what might have caused it (e.g being killed by robbers, natural death, due to reckless lifestyle and due to accident). Although the majority of respondents concurred that it was all right to talk about death, they put conditions and suggestions on how it should be talked about and with whom. There were suggestions that it is important to specify the type and cause of death being referred to. The general feeling was that death is a natural, normal and inevitable event in one's life and that everybody expected to die at some point in their lives. Hence referring to natural death was regarded as acceptable. However, the local term for 'dead' (*gukua)* was regarded as evoking strong and or frightful feelings and a variety of softer and age-group specific terms with the same meaning as dead were suggested. These were terms such as *gwitwo* (to be called), *kwehera thi* (move from the earth) , *guthii* (to go), *kuinuka* (to go home), *rugendo gukura* (the journey has matured or literally "has become old"), *kuhuruka* (to rest) among others. An interesting and more light-hearted reference to death amongst the young was that "*someone has timed a trip to heaven"* which has similar connotative meaning with the other terms that the dead had gone to another place.

With respect to SG, 50% of the respondents said they were willing to risk death especially if the condition is very serious and the problems were too much, noting that there are

conditions that are worse than death while others simply had no problems as they recognized they can die any time.

Thirty eight percent in both VAS and TTO felt that death was mentioned too often and it evoked strong reactions. Respondents believed that death was not represented as a serious issue in valuation tools. It was annoying for those recently bereaved or facing death in the family. Death was directly associated with beliefs that talking about it actually brings it towards you. Respondents noted that some people might get the wrong perception and think that if they imagine themselves being dead they will die soon or, that the interviewer had seen they were about to die that is why they ask them about death and also that everyone fears death. Others may misunderstand your motive and take it badly. It was also mentioned that it would not be a good idea to ask the very sick about death (6.25%).

Twenty five percent on whom SG was pre-tested were unwilling to risk immediate death and attributed their opinion to the fact that this was like loosing hope in life, the fact that one can never be in perfect health and fear that one may start thinking that they are sick and may die. Another 31.25% were non-committal saying they would gamble depending on the nature of problems and the age of the person. They noted if the problems were too much they would gamble and also that the young may be unwilling to risk death hoping to get better and accomplish some things in life, while the old have already accomplished and would rather "rest once and for all" rather than live in problems.

Imagining or thinking of being very sick: Asked whether it would be bad to ask someone to imagine themselves being in a very sick state of health, the majority (94%) responded no. One respondent mentioned that it would be better to ask the person to think of another person as it is not good to ask someone to think of him or herself as being sick. However, all respondents agreed that asking about a sickly health state would not cause sickness in the community.

7.3.5.4.2 Time trade off

Giving up time: Asked about their opinions about being asked to give up time in bad health to stay in good health, the majority (81.25%) said they would be willing to trade off time, 6.25% was unwilling to trade off and 12.5% were non-committal. The unwilling respondent said that she was arthritic but would prefer to live in her condition rather than trade off time to avoid the pain. Those willing to trade off time cited reasons such as: to avoid depleting family resources when sick; being unable to fend for themselves; and hating to live in illness. Other reasons included love for good things (perfect health), ability to utilize time productively when healthy and not being a liability.

7.3.5.4.3 Standard gamble

Gambling with perfect health and immediate death: There were mixed reactions to being asked to gamble with two states of health. Although the majority (56.25%) would be willing to gamble to avoid living in problems, they would only gamble if the state was too bad and also the realization that gambling does not mean that you would die. A few (18.75%) were unwilling to gamble due to fear of gambling should the worst outcome result and conviction that they could never take a chance with death. One among these said that, *"if it was real and you ask me to gamble, I will not gamble because you do not know what will happen"*.

Comparing certain and uncertain life: Presented with the prospect of choosing between a certain and an uncertain life, 37.5% said that they would be in a dilemma over which one to choose. This was because in the uncertain one perfect health is good but immediate death is bad and in choosing it you would be in doubts, while the certain one has problems. 31.5% said they would not feel anything while 6.25% said they gamble and hope to live. 12.5% did not like imagining the possibility of death while 12.5% thought that perfect health was unrealistic.

Comfort in dealing with probability: Respondents were asked how comfortable they were in dealing with probabilities. Majority (87.5%) said they were comfortable since life is about chances and possibilities. Only 12.5% said that they were uncomfortable with the idea of probability of immediate death, because they considered themselves "unlucky" when it comes to gambles.

7.4 DISCUSSION

This chapter presented the methods followed in the construction of Schistosomiasis Mansoni disease states and assessment of content validity of three valuation instruments. Content for disease states were ascertained during the development, subsequent use and construct validation of the schistosomiasis HRQL questionnaire addressed in chapters five and six. Content validity for the disease states is therefore not discussed here. This discussion focuses on the content validity of the VAS, TTO and SG in the Kenyan setting. The discussion attempts to address three issues namely: In what ways were aspects of content validity assessed in this study similar and different to the general approaches for assessing content validity? Can the VAS, TTO and SG be considered content valid instruments for eliciting values for disease states in Kenya? And, should any of these instruments be used with the Kenyan population?

Validation of any scale should start with ascertaining its content validity to avoid the prospect of a scale that performs well on other psychometric criteria or otherwise, but measures what it was not intended to measure. A content valid scale should cover the domains under investigation adequately (Streiner and Norman, 1995). Approaches that have been used in content validation of health measurement instruments have mainly included: examining literature; gathering expert's opinions (i.e. talking to patients and professional) to chose the domains that should go into and instrument (Trudel, et al. 1998, McDowell and Newell, 1996); as well as opinions from the lay population for whom a measure is intended (McDowell and Newell, 1996; Hunt et al., 1991). Further to including lay concerns, Hunt et al., (1991) suggest that while using instruments cross-culturally, it is the conceptual rather than the semantic equivalence that should be pursued.

Studies on content validation of health state valuation instruments are lacking. Recently, Brazier and Deverill (1999) and Brazier et al. (1999b) have provided a checklist of for judging the merits of preference based measures, (which includes validity) and applied these criteria to a review of valuation instruments in Brazier et al. (1999b. P.23-56). The criteria provided addresses validity from a theoretical and empirical perspective. Therefore, evidence on content validity of valuation instruments is still lacking. It is important to note that there are no set criteria for claiming existence or lack of content validity (Cano, 2001; Streiner and Norman, 1995; McDowell and Newell, 1996). Claims for content validity typically rest on the comprehensiveness of the instrument and the methods used in constructing it (Brazier and Deverill, 1999). In general one examines the content of items comprising a questionnaire and judges their content for relevance to the task at hand in determining that the instrument measures what it intends to measure. The items are also assessed in terms of whether they would be acceptable, appropriate, understood by the respondents (Brazier et al. 1999). This study has assessed content validity of the VAS, TTO and SG using a qualitative approach to determine the existence of concepts embodied in the instruments and whether respondents can and are willing to contemplate the tasks involved in each valuation task.

While underscoring the importance of other forms of validity relevant to valuation instruments (Brazier et al. 1999), assessment of content validity of the instruments incorporating lay concerns was considered an important initial step to transferring the instruments to a new (Kenyan) setting. However, an important step in cross-cultural transfer that was omitted in this study would have been to contact the instrument developers to document what domains of content and concepts they considered central to the instruments for the populations they were used in, before investigating their content validity and equivalence (Herdman et al. 1998) in Kenya. This would have offered an opportunity to assess and compare the conceptual equivalence of the methods in the two settings. Nonetheless, except for TTO (Torrance et al. 1972), it is not clear from the literature who could be contacted for the development of the SG (attributed to von-Neumann and Morgenstern) and VAS. Despite this shortcoming, the author contends that the selection of

terms and concepts embodied in each instrument is unlikely to be different because the questions constituting the process of valuation may be thought of as the items upon which selection of terms and concepts was based. By examining whether the terms and concepts embodied in the instruments existed, assessment of content validity of valuation instruments sought to determine if the instruments would be understood, accepted, found appropriate and practical in this community. What were the respondents' views and opinion regarding the whole idea of the valuation process? It was envisaged that understanding, accepting and deeming the process of valuation appropriate or otherwise, would have an effect on the values obtained therefrom, thereby impacting on the validity of such values. If the process of obtaining values is faulty (because terms and concepts in the instruments are not well understood, or are inappropriate or the process is generally unacceptable for various reasons), then it would be incorrect to claim validity of the values obtained from such a process.

The methods adopted in this study to assess content validity of the instruments is considered novel considering the paucity of studies assessing content validity of valuation tools. However, the method has its strengths and weaknesses. Although the assessment of content validity of the three instruments by the same respondents took long and therefore required a lot of commitment from the respondent, it allowed comparisons and therefore generated more useful information. The debriefing interviews held immediately after the completion of the exercise allowed the respondent to reflect on various aspects of the use of the tools. One to one interviews also allowed for independent views from each respondent as opposed to use of focus groups where some vocal people might dominate. Another possible weakness was that, as the interviews did not restrict the responses to health related issues, the responses may have been too general and difficult to relate back to health state valuation. Therefore in future, it would be helpful to ask the respondents to think of valuation instruments in relation to valuation of health related interventions, as perhaps the responses might be different. Considering the length of the interviews, perhaps a better strategy would have been to use focus group discussions to evaluate the tools as the groups have more members and therefore likely to elicit diverse views. Also collecting views from a larger sample would cast more light especially on those issues that were found to present

problems in the instruments. Finally, if time allows, participant observation of how people make trades, what and how they value different aspects, along with observations of other aspects embodied in the instruments would provide useful information into content validity of instruments.

Can the VAS, TTO and SG be considered to have content validity for eliciting values for disease states in Kenya? The VAS, TTO and SG have concepts that are both common and specific. Common concepts like the ability to visualise or imagine oneself in a scenario as well as the concept of comparisons were found to exist and were understood as judged by examples and instances provided, where members of the community encounter them. However, although the concept of perfect health was well understood and described by the majority of respondents, it was found impractical, unrealistic and unimaginable by about 18% of the respondents, who found it difficult to think of such a state. This implies that the best health state, with no problems was rated well below 1.0. As thc best state is forced to be given a value of 1.0 in the TTO, and the SG, it would imply some re-scaling of value is needed before valid comparisons can be made across cultures. Inability to conceptualise perfect health may, in my view, influence how people perceive these techniques as realistic instruments for valuing health. In addition, inability to perceive perfect health implies that though people can use the techniques, the values so obtained may not really represent what the instrument developers intended. Also, the fact that people are unable to imagine perfect health means that they may consider a valuation exercise as an exercise in futility and not give it much thought.

All respondents provided instances and examples of concepts specific to the VAS and TTO (such as ranking, expressing preferences, choice, giving up, future, time and life). It can therefore be argued that assuming people have knowledge of what they are choosing between (through imagining) they can express their ordering of preferences by ranking using VAS. For the TTO, concepts like giving up and trading off exist as evidenced by existence of local words expressing them. The concept of future was found to exist and its expression involved uncertainty. In thinking of the future people appeared more comfortable with the near future, specifically one year, which coincides with farming

seasons characterized by given months. They could also relate to time both in terms of duration as well as by units such as hours in a day, months and year. In referring to the concept of life, it was seen as a totality of ones achievements and endowments while living, as these determine the activities one is engaged in and how these shape who one is. This then, is perhaps what people take into consideration while comparing two lives and deciding which is better than the other.

In the SG, three respondents failed to provide examples and instances of where concepts of gambling and uncertainty were found, while one respondent failed to provide examples of concept of risk and probability. This might be an indication that these respondents had difficulties conceptualising these concepts and may therefore experience difficulties using the SG. Although, for the majority, the concepts of risk, probability or chance, gambling and uncertainty were well understood and easy to relate to in this community, asking people to gamble with death may become unacceptable especially in real life situations. There was indication of unwillingness to gamble with comments like, "*if the chance of death were as high as 1%, I would not gamble*". Hence, there are doubts about whether application of the SG in a real situation (say in a hospital setting with actual patients being asked to take a risky procedure that could cause immediate death, but with a chance of recovery) in this community would be feasible. This is due to the fact that the technique relies on the willingness of the respondent to trade off risk in order to improve their health (Brazier et al. 1999).

Tasks like placing a numerical value on a health state, imagining or thinking about death and being very sick are common to all instruments. The majority had no problems with placing a numerical value on a disease state although a quarter of the respondents found it difficult. Asking people to think of being very sick was not considered bad. Death appeared to be a major issue in all instruments. Views about death appear contradictory in that majority did not find it a problem to talk about death while minority hold opposite views. The idea of asking people to think about death, risk death was cited as a problem with regard to all instruments and more so with the TTO and SG where mention of death was often. This could be considered to pose more serious problem with the SG and TTO

because in valuing chronic states and states worse than death the mention of death is almost inevitable while with the VAS, one could replace death with the worst imaginable state as in EQ-5D VAS. However, if death is one of the states being valued, it is still inevitable to mention it. For the TTO, depending on the duration of the state being valued, it is possible to replace the end point with "what happens after is not known" thereby circumventing the problem of mentioning death too often.

Other ways of circumventing the problems attributed to mentioning death would be to use softer terms that do not arouse negative and unpleasant feelings and generally clarifying what type of death you are referring to. Talking about 'natural' death (i.e. not caused by violence or reckless living, but normally, of old age) was deemed more acceptable. This raises the question of whether a young person can die a natural death and underscores how complex the issue of death can get. It was less clear whether dying of ill health at a young age would be considered 'natural' and whether this would be acceptable. As a natural thing, no one would be completely at ease talking about death because it is an unpleasant event that many do not desire. However, this intrinsic characteristic of death does not preclude its mention, despite the unpleasant emotions it provokes in people, and one should be informed, sensitive, careful and cautious in engaging in death talk in different communities. It appears that the way the issue of death in instruments is handled can determine the type of responses one gets. Good translation can never get round this issue and instruments need to be changed to account for locally appropriate approaches to 'death talk'. It would therefore be useful to find out before hand how people in a community generally refer to death in terms that are acceptable and do not arouse strong negative feelings. Research into how different types of death would affect values for disease states is vital in cross population comparisons of values. For example would the value for disease state j be different if it was followed by 'natural' death or 'unnatural' death, even though it was the disease state and not death that was being valued?

That almost 19% were either unwilling to trade off time or were non-committal signals that use of TTO may encounter problems. This is because the technique relies on the willingness of the respondent to trade off life years in order to improve their health and

consumer theory assumes that individuals will trade to maximise utility (Brazier et al. 1999). Unwillingness to make any trade of health for probability or quantity of life raises serious questions about not only the presence of lexicographic preferences, but also whether this is likely to differ by culture – or possibly by perceived risk of mortality. With respect to gambling with perfect health and immediate death, one respondent remarked, "*if it was real and you ask me to gamble, I will not gamble because you don't know what will happen*", casting doubts how useful SG would be in a real situation.

The prospect of choosing between a certain and uncertain life elicited responses suggestive of peoples attitudes to risk. Some were in dilemma (risk *cautious*), others said they would not feel anything (risk neutral), others would gamble and hope to live (risk loving) while others hated the possibility of death (risk averse). That these are so diverse casts doubts on theoretical and empirical validity (Brazier et al. 1999) of SG. For the risk *cautious* group their feelings of being in a dilemma points to aspects of regret and rejoicing as found in Regret theory (Loomes and Sugden, 1982).

The majority judged all the techniques as appropriate and suggested possible changes regarding those aspects considered inappropriate. Mentioning death too often was one of the aspects considered inappropriate. There were difficulties in visualising scenarios related to the amount of information contained in a scenario, suggesting that the scenarios needed revision with a view to reducing the information. The fact that people would not be comfortable with thinking that the states being described are their own, and would worry that it was their states being valued was a cause for concern regarding practicality in having someone have knowledge of and valuing their own state. However, people in this community appeared comfortable in valuing states belonging to "others" not known to them rather than their own. This has implication for issues of whose values to use. Would patients, for example, be conformable with contending with the states they are in and valuing them? These concerns were accompanied by suggestions on how to improve the instruments. They included; avoiding the frequent mention of death by replacing 'and then you die' with "what happens after is not known", emphasising that the states being valued are not in any way those of the respondent, to avoid unpleasant feelings and sentiments;

assigning names to disease states and emphasising from the outset that they are not those respondents are in. These suggestions were incorporated in making changes to the VAS and the TTO and are further discussed in chapter 8.

Although respondents were able to make some differentiation between scenarios, it required the scenario being described twice, and even then not all the differences were discerned. This may imply that, in assigning a value to a disease state, not all pieces of information are utilised which may raise concerns regarding interpretation and comparison of values for different individuals and similar states. Are all respondents assigning a value to the same state or to parts of the state? Are these values comparable and what do they represent? Answers to these questions deserve further investigation. This finding also draws attention to assessment of practicality of instruments in that what the respondent states in a dichotomous response type of question (i.e. do you understand the task? Yes, No), is not necessarily what is found in practice. It is therefore important to ascertain practicality by putting respondents through tasks that indirectly inform on whether they understand the tasks or not.

It was mentioned that the very old, children, the very sick and illiterate may have difficulties using the techniques and that they might provide unreliable responses. Researchers would therefore need to take these groups into account when using these techniques with them, but how the techniques could be modified to suit them is an issue for further research. Also needing further study is the extent to which these groups would be (un)able to use these techniques and what the critical age would be on either end.

Based on the above discussion, it is noted that no instrument was problem free. Going by majority rule, all the instruments are deemed to be content valid to a reasonable extent for use in Kenya, but taking into consideration the concerns raised for each instrument. However, a significant minority had problems conceptualising perfect health, dealing with mention of death and placing a numerical value on a disease state. The SG and the TTO also had a few respondents expressing unwillingness to gamble or trade off time. With respect to SG there was also the issue of consideration of whether the SG task was for

actual decision-making or just a hypothetical exercise, with the view that if it was for real people would not gamble. While these concerns are held by few they cannot be ignored because they are central to the functioning of these instruments and deserve further research as they have implications on theoretical and empirical validity of these techniques. However, this being the first time using these techniques in this area, it would be premature to state categorically that the instruments would work or not and this calls for more studies on content validity of these instruments. However, the issues raised especially with regard to conceptualisation of perfect health, death talk and people's willingness to trade probability or quantity of life lead to questioning the validity of comparisons of values across cultures as a route to determining cross-cultural comparability. Concepts in existing valuation instruments can be perceived very differently. However, assessing conceptual equivalence can help consider potential and necessary modifications of the instruments and improve relevance to local cultures and understand the extent to which comparisons of values across countries is meaningful.

Due to time and financial constraints and sensitivity to respondent burden, further research on validity and reliability of the valuation instruments was undertaken using two instruments. Choice of these two instruments was based on which instruments performed best on issues of content validity considered. The three instruments were ranked 1,2,3 where 1 was best performing. Where there were ties the score was added up and divide by the number of ties so that the row score total was six. This ranking was based on proportion of respondents for example expressing an opinion. For example, on the issue of death the technique with most respondents expressing concerns or having problems is ranked as performing worst (3). For the four aspects assessed, sub-totals and the grand total of these ranks were compared for three instruments. The instrument with the lowest score was considered the best. Table 7.12 shows the summary of the ranking that was obtained by summarising the results on each aspect.

Table 7.12: Summary of performance of the VAS, TTO and SG: Comparisons on four aspects on content validation (Ranks: 1=best and 3=worst)

Aspect being assessed	VAS	TTO	SG
Existence of terms and concepts			
Instrument specific terms and concepts	1.5	1.5	3
Common concepts	2	2	2
Sub-total: Existence of terms and concepts total score	***3.5***	***3.5***	***5***
Ease of understanding and use			
% understanding task	2	2	2
No of times task explained	1.5	1.5	3
Ease of understanding	1	2	3
% experiencing difficulties	1	3	2
Technique recommended	1	2	3
Sub-total: Ease of understanding and use total score	***6.5***	***10.5***	***13***
Appropriateness of techniques			
Annoying	2	3	1
Embarrassing	2.5	1	2.5
Convey wrong meaning	1.5	3	1.5
Sound disrespectful	2	2	2
Cause problems	1	3	2
Offend or upset	1.5	3	1.5
Sub-total: Appropriateness total score	***10.5***	***15***	***10.5***
Practicality in use of valuation techniques			
Extent of thinking	1.5	3	1.5
Things respondent worried about	1	3	2
Feelings thought or emotions provoked	2	1	3
% expressing worries, thoughts and feelings	1	3	2
People to whom technique should not be administered	2	3	1
Instrument specific tasks			
Tasks common to all instruments			
Placing a value on a health state	2	2	2
Death	2	1	3
Tasks specific to instrument			
Giving up time	na	6	na
Gambling	na	na	6
Dealing with uncertainty and probability	na	na	6
Sub-total: Practicality total score	***11.5***	***22***	***26.5***
Grand total score	***32***	***51.0***	***55.0***

The summary shows that VAS performs best on all aspects considered, indicating that it would be more acceptable, practical and appropriate and would not present major problems with conceptualisation. This instrument should therefore be tested further in Kenya, as this would also provide additional information to findings of Kirigia (1994). The TTO performs better than the SG on all aspects except appropriateness, a problem largely attributed to mentioning death too often, thereby slightly outdoing the SG. This finding is common in other studies (Dolan et al. 1996b) that find the TTO and SG performing more or less the same. Although both TTO and SG have problems with some concepts and tasks that are central to their use, choice of TTO for further testing in Kenya was based on the its better performance especially on terms and concepts, ease of use and understanding and practicality. However, further research should consider the use of the SG, as it is a choice based method with the strongest theoretical foundation and it yields utilities rather than values.

7.5 CONCLUSION

Although the sample was selected conveniently, the age range covers young adults to the old, the majority of whom are married and engaged in farming. This implies a fair representation of these groups' views in the findings, as their characteristics are relatively similar to those of the community (chapter 6).

This chapter set out to present methods for the construction of disease states and approaches in the development and choice of valuation techniques for use in Kenya. Six disease states were constructed, some of which were used in the pre-testing of valuation techniques. From the findings, we conclude that there is need to reduce the information contained in the disease states. Also inclusion of frequency (in terms of days) in expressions of the descriptors of disease states was found to be repetitive, confusing to respondents and akin to double counting time especially because the states were described

to last for a year. This suggests that in describing the disease states frequency of symptoms be omitted and represented by same duration of one year for all states[60].

Personal observations regarding the procedures for training interviewers and pre-testing revealed that the exercise is time consuming and requires extensive explanations of why, how and for what reasons these techniques are used. Given that this sort of activity was relatively new in this community, respondents considered this as an educative exercise because they were learning something new. Therefore application of these techniques for the first time in a community may be taken more as a learning venture on how the methods are used. This may imply that with repeated use the performance of the methods may even be better.

Based on assessment of existence of terms and concepts embodied in the valuation techniques, most of the terms form the three instruments were found to exist. However, terms like death, gambling an dealing with uncertainty, trading off time and imagining perfect health presented some difficulties and therefore require careful consideration if instruments are to be used in this community. Judging by ease of understanding and use of the techniques, VAS and TTO were found to be the most preferred techniques. All techniques were considered appropriate although there were more problems associated with TTO and SG. A summary of performance of the three techniques scored VAS best followed by the TTO and SG, where the TTO performed better than SG in all aspects except appropriateness. In general the techniques are considered to attain a considerable degree of content validity in this setting. The VAS and TTO were chosen for use amongst people of Mwea. Use of these two techniques amongst patients and community members and further testing of their performance in terms of construct validity and reliability is dealt with in the next chapter, together with the changes made to the valuation instruments.

[60] Changes made to the disease states and valuation instruments are presented in chapter 8.

CHAPTER 8

VALIDITY AND RELIABILITY IN VALUING DISEASE STATES: THE CASE OF SCHISTOSOMIASIS IN KENYA

8.0 INTRODUCTION

Evidence from assessment of content validity of the VAS, TTO and SG in Mwea lead to selection of the VAS and TTO for further testing of performance of these two techniques using schistosomiasis disease states. Chapter 4 established that assessment of disease specific utilities is a fairly unexplored area especially in developing countries and also for tropical conditions like schistosomiasis that afflicts the majority of populations. In addition, the few studies using valuation instruments to assess disease specific utilities in settings other than those in which they were developed recognise the existence of cross-cultural differences (Ashby et al. 1994; Kirigia, 1998 and Baltussen et al. 2002). Furthermore, the choice of valuation instruments has largely been guided by their popularity and ease of use rather than their validity and reliability for specific settings, as established in chapter 4. Ensuring validity, reliability and practicality of instruments in new settings enhances their credibility as well as the data they generate. This underscores the importance of ascertaining the methodological successes or failures of the techniques in terms of practicality, validity and reliability before they can be usefully employed for measurement of disease specific utilities (DSU) in the new setting for health care decision making.

This chapter aims to determine the practicality, validity and reliability of VAS and TTO in valuing disease (health) states using schistosomiasis as a case study. It also aims to gain an understanding of the factors influencing variation in values for the same and different disease (health) states.

In the next section, methods for assessing practicality, validity and reliability are presented together with the sample selection, instruments and props and a description of disease states

used for the study. Section 8.2 presents the results followed by discussion and conclusion in sections 8.3 and 8.4 respectively.

8.1 METHODS FOR ASSESSING PRACTICALITY, VALIDITY AND RELIABILITY

8.1.1 Samples and sample selection

Four sub-samples were used in this study: a patient and community sample and a test re-retest sample for each. The sampling frame consisted of the measurement study sample described in chapter 6, i.e. 80 patients and 81 community members.

To select the patient sample for the initial valuation exercise, the 80 patients who had consented to be followed up for this phase of the study were assigned numbers from 1-80. Using Microsoft Excel random numbers were generated from which a sample of 60 patients was randomly selected[61]. To select the test re-test sample, a similar procedure was followed using the initial valuation sample to randomly select 30 patients.

The community sample used the same method as for the patients. However, out of the 60 randomly selected community members from a total of 81, there were three refusals, thereby reducing the initial community valuation sample to 57, from which 30 were randomly selected for the test re-test exercise.

The time lapse between measurement and valuation study was nearly a year. At the time of valuation, the disease status of those previously ascertained to have schistosomiasis (referred to as patients) was not determined. Therefore, the patient and community sample were not considered to be different with respect to disease status, except that those in the patient sample had had schistosomiasis mansoni and had been treated a year earlier. Therefore, analysis of validity and test re-test reliability was based on the pooled sample of

[61] Note that this sample is not purely random because the sampling frame upon which it was selected was a purposeful and consecutively selected sample (see chapter 6).

117 and 60 respondents respectively. However, the results on characteristics of respondents and practicality of the instruments are reported for the patient and community group separately so as to make group comparisons. The decision to sample 117 and 60 respondents for initial and retest valuation was based on time and financial considerations. Also, the purpose of the study was to test the methods and therefore this sample size was considered sufficient to allow for statistical testing.

8.1.2 Instruments and props

Two instruments namely, a valuation script and a record form were used. The valuation script (appendix 8.1) contained the instructions and steps followed for both the VAS and TTO valuation tasks as well as the specific questions posed to the respondent to facilitate valuation. For the TTO there was a script for states better and worse than death. In line with the concerns raised pertaining to mentioning death too often during the pre-testing of TTO, the question for eliciting the value was revised and was posed as for example, *"Now I have changed life A scale to 2 months* (point to the pink card). *Life B scale remains unchanged at 12 months* (point to the appropriate coloured card). *This means that you would either live in life A like* (mention name of person on pink card) *for 2 months or live in life B like* (mention name of person on appropriate coloured card) *for 12 months. What happens after is not known. Would you prefer life A or life B or are they equal?"* This avoided the frequent mention of the fact that after a given duration in perfect health, one dies.

In addition to collecting information on respondent characteristics, illness and health status, the record form (appendix 8.1) also contained records of values for different disease states. Other areas addressed in the form were ease of use and understanding and practicality of the TTO and VAS. Both the valuation script and record forms were translated into the local language, Kikuyu, by the research team[62] before they were administered.

The props included a VAS board and TTO board (appendix 8.2) constructed following Furlong et al. 1990's guide to design and development of health-state utility

[62] This followed a similar procedure as outlined in chapter 7 section 7.2.3.

instrumentation and six colour-coded disease states (appendix 8.3). The VAS board was a straight line with endpoints marked 0 (worst imaginable health state) and graduated in intervals of 10 to 100 (best imaginable health state) and had several pointers for indicating the value assigned to a given disease state. The TTO board for states better than death was 41 by 56 centimetres board divided into two portions representing two lives. Each portion had a pocket for holding scenarios being valued and a scale in months running from 0 to 12 months. The scale for the top portion representing life in perfect health was pink while that for the lower portion representing the life in a state being valued was green. There was a slight variation in the TTO board for states worse than death. The top portion represented both the state perfect health and the disease state being valued and therefore had two pockets. The lower portion had one pocket containing the state 'dead' and therefore had no scale.

Four disease states were used in the valuation exercise with state 'perfect health' and 'dead' as endpoints. A separate disease state was also used in the mock exercise before the actual valuation started to demonstrate to the respondent how the instruments were used. The four disease states represented mild, moderate, severe and very severe disease states. They were colour coded differently, assigned imaginary names (male and female) as follows and laminated:

Table 8.1: Disease states identification

Identification letter	Disease state level	Imaginary name	Colour code
PH	Perfect health	Paul / Peris	Pink
A	Mild	Anthony / Anne	Purple
C	Moderate	Christopher / Carol	Light blue
E	Severe	Eric / Emma	Yellow
F	Very severe	Francis / Faith	Orange
ZZ	Dead	Zachariah/ Zipporah	Blue
MM	Moderately severe	Mock state	Green

8.1.3 Description of disease states to be valued

During the pre-test (chapter 7), a number of issues regarding the disease states became apparent and necessitated several changes in the scenario description. These issues related to: number of scenarios being valued and the time factor; information overload; specification of the duration a state lasted; lack of differentiation of scenarios as they were all colour-coded green and had no identification names.

Initial attempts to value 8 states resulted in interviews lasting over 2 hours which was considered too large a burden on the respondent and could compromise the quality of the responses due to respondent fatigue. An arbitrary decision to value only four disease states (Appendix 8.3) using TTO and 6[63] states using VAS was made to shorten the total interview duration.

Information overload in disease state scenarios was cited as a problem during pre-test. To reduce information in the scenarios, symptoms were retained in their original form but the number of HRQL domains included was reduced to three namely; performance of work and daily duties, social participation and feelings of worry and anxiety. Their choice was justified on the basis of commonly considered dimensions and noted differences in scenarios during pre-test, which included ability to work and socialize. Hence, within the physical domain, the dimensions of mobility , energy and strength and performance of output and work were excluded. This was based on the argument that if one was able to work and socialize, then one was presumed to have energy and strength and to be able to move about (mobility). However, this argument could be challenged on the basis of not entirely reflecting the results from the measurement study that supported the importance of feeling of strength and energy and mobility[64]. The four disease states therefore contained the following symptoms along with the three HRQL domains mentioned above: tiredness, loss of appetite, watery and bloody mucoid diarrhoea and itching skin rash. These scenarios

[63] In the VAS, the states perfect health and dead, which are not ordinarily valued using TTO, were valued, hence the six states.
[64] However, the three domains that were included also reflect WHO's (1993) conception of health.

were described in a manner that would make the differences distinct and is summarized in table 8.2.

To cater for uniformity in the duration the disease states lasted and to avoid mentioning time in terms of differing days which caused confusion, the expressions of frequency in terms of days in the scenarios was removed. All scenarios were specified to last for duration of one year, which respondents would associate with as it coincides with rice farming activities.

Table 8.2: Description of disease states

Symptoms and HRQL descriptors	Mild disease state A	Moderate disease state C	Severe disease state E	Very severe disease state F
Tiredness	Somewhat	Somewhat and then very	Very	Extremely
Loss of appetite	Can eat a ½ to a ¼ the normal amount of food	Can eat a ½ to a ¼ and then no more than 2 spoonfuls the normal amount of food	Can eat a ½ to a ¼ and then no more than 2 spoonfuls the normal amount of food	no more than 2 spoonfuls the normal amount of food
Watery and bloody diarrhoea	Nil	Nil	Sometimes	Most of the time
Itching skin rash	Nil	Nil	Moderate	Moderate
HRQL domains: performance of work and daily duties; social participation; worry and anxiety	A little of the time	Some of the time	Most of the time	Most of the time

Colour-coding all disease states green made them seemingly appear the same to the respondents and caused them confusion when asked to compare the states. As a result, the six states were coloured differently and assigned imaginary names as spelt out in table 8.1. The assignment of names was also meant to clarify that it was not the respondent's disease state being valued but that the respondent was valuing a disease state of another person. However, they were asked to imagine that this person was just like them.

8.1.4 Computation of VAS and TTO values

The VAS values were all positive as no state was considered worse than death using the method. Hence, there was no need to transform the values. However, VAS values were presented scaled to a 0-1 scale by dividing by 100 to be comparable with the TTO values. The TTO values for states better than death and worse than death were computed as $h_i=x/t$ and $h_i= x/(x-t)$ (Torrance, 1986) respectively, where h_i is the value for the state being valued, x is the duration in perfect health and t the duration in state being valued. For the positive values there is usually no need for transformation. However, for the negative values there is no theoretical lower boundary on the scores for states worse than death (Patrick et al, 1994) and this produces asymmetry between positive and negative ends of the preference scale (Badia, et al. 2001). For example, our TTO scale was 12 months, producing a lower boundary of -23, a very high value that would affect correlations and means. To make the negative and positive side of the scale symmetric i.e. -1 to 1 scale, the values for states considered worse than death were computed as $h_i=x/(x-t)$, and then transformed using the formula, $h_i/(1-h_i)$ (Patrick et al. 1994; Badia et al. 2001) to a -1 to 1 scale.

8.1.5 Analytical methods

8.1.5.1 Practicality

Practicality was assessed in several ways namely: ease of understanding and use of techniques as judged by the interviewer and the respondent; thinking effort required to complete the valuation task; time; and types of people who could use the technique. Assessment of each is described fully below.

Ease of use and understanding of technique was assessed both from the respondent's and interviewer's perspective. Respondents were taken through an example of valuation using a

separate disease state and the number of times this was repeated was recorded to indicate how easy or difficult it was to understand use of the technique. Ease of understanding was assessed by rating on a 5-point scale from 'very easy' to 'very difficult' as well as pointing out the difficulties encountered during use of each technique. The two techniques were then compared in terms of ease of understanding the questions and the exercise in general as well as indicating the preferred technique and why.

After using each valuation technique, the interviewer assessed the exercise in general based on three aspects. These were: a rating of respondent's level of difficult in understanding the valuation task on a 5-point scale from 'not difficult at all' to 'very difficult'; a rating of respondent's level of comfort with the way the questions were posed, on a 5-point scale from 'not comfortable at all' to 'very comfortable'; and a rating of the whole exercise and the responses that were obtained in general, on a 5-point scale from 'very good' to 'very poor'.

Respondents were asked to indicate the extent of thinking they required to use each technique on a 4-point scale from 'very hard' to 'very little'. It was assumed that if respondents required thinking very hard, then the task would be exerting more mental burden and exhaustion on the respondent, which may lead to unacceptability of the method in the community[65].

The time taken to complete valuation using each technique was also recorded in minutes. For the VAS, this was time taken to value six states while for the TTO it was time taken to value four disease states. Respondents were also asked to state their opinion about all the different types of people who should not be asked to undertake valuation task using each method and why.

[65] The opposite view could also be argued that if people don't think hard about their choices then they are not giving the exercise the importance it deserves, by not giving it much thought. However, other aspects of assessment of practicality (as shown in results section) support our assumption.

8.1.5.2 Validity

Content validity was assessed in chapter 7 and argued to exist to a sufficient extent for TTO and VAS to warrant further testing. This will not therefore be addressed further. Criterion validity was not assessed, as there was no criterion against which to judge the validity of VAS and TTO. Construct validity was therefore the only form of validity assessed and tested as described below.

8.1.5.2.1 Construct validity

Construct validation is about hypothesis testing concerning specified relationships between the constructs and the use the test is being put to. In this case, the construct was that the disease states represented less desirable states of HRQL and that more severe disease states would be less desirable hence receiving lower values. Hypotheses that were tested are specified below along with the test methods used.

Given the similarity of the disease states valued, it would be expected that different valuation techniques produce similar values for similar disease states. However, this expectation is not supported by the literature (Torrance, 1986; Dolan et al. 1996) and therefore it was not possible to state a priori the expected behaviour in terms of similarity of values for a given state from VAS and TTO. Therefore, a null hypothesis is posed that there are no statistically significant differences in mean values obtained using VAS and TTO for the four disease states. This hypothesis is tested using the *t-test* for differences between means as well as the Wilcoxon sign rank test.

In as far as the valuation techniques are valuing similar constructs, VAS and TTO values for each of the four disease states were hypothesised to have positive correlations. The Spearman's correlation coefficient was used to test the correlations between VAS and TTO values for each state.

Given that the disease states reflect worsening severity, it was hypothesised that the worse a state was, the lower its value would be and vice-versa. This was expected to hold for both VAS and TTO valuations. Mean values of disease states from both techniques were compared in terms of magnitude to explore whether worse off disease states were assigned lower values. It was also hypothesised that ranking of health states using both techniques, according to mean values, would reflect the logical ordering of the disease states in terms of severity, to reflect the ranking, A>C>E>F.

Further assessment of construct validity considered the extent of 'logical inconsistency' in ranking of disease states. Assessment of inconsistent ranking was done by recording the number of disease states inconsistently ranked with disease severity and the proportion of respondents ranking a given disease state inconsistently with disease severity. To assess inconsistency in ranking, it was assumed that states would follow a logical ordering according to severity of symptoms and extent of disruption of HRQL domains. Disease states had thus been chosen to reflect mild (A), moderate (C), severe (E) and very severe (F) disease states with perfect health (PH) and dead (ZZ) as end points. A logical ordering was thus expected to rank the disease states as: PH, A, C, E, F, ZZ. After valuing the states using each technique, they were ranked according to the values assigned and those not corresponding with their logical ordering positions were considered inconsistently ranked. For example, in VAS, PH, A, C, E, F and ZZ should have been logically ordered as 1,2,3,4,5,6. If a respondent's values ranked the states as 1,3,2,4,5,6, then only four states PH, E, F and ZZ would be considered as logically ranked and states A and C would be considered as inconsistently ranked. A similar procedure was followed for TTO.

Examination of factors affecting values was considered as an aspect of construct validity. Consumer theory points to a number of factors that may affect values, amongst which are; human characteristics, attitudes and behaviour (rationality, consistency and transitivity), socio-economic and demographic factors, characteristics and amount of information about the good, consequences of the choice, whether certainty or uncertainty prevails during the choice process. This study examined how both VAS and TTO values varied by gender, age, marital status, whether respondent had children or not, number of children, education level,

monthly income and expenditure. Also assessed was whether illness during valuation, experience with schistosomiasis and health status (usual and in the two weeks prior to valuation) caused variation in values. Reviews in chapters three and four indicated that there is no agreement on how values vary by age, gender and educational level. Hence a null hypothesis of no difference in mean values between groups with respect to these factors was tested. Although the majority of studies (chapters 3 and 4) indicate that those in ill health, have had experience with disease or are in poor health assign higher values to health/disease states, a few studies (Kirigia, 1998; Chen et al. 1996; Dolan 1996) have found negligible or no relationship. Therefore, a null hypothesis of no difference in mean values for disease states between those experiencing illness during valuation, those who had experience of schistosomiasis mansoni, those rating their usual health status and health status in the two weeks prior to the valuation low and those not was tested. Marital status and whether one had children or not had not been previously investigated and therefore a null hypothesis of no difference in mean values between groups was tested. From consumer theory, positive or negative income effect (and proxy expenditure) depends on whether the good is considered normal or inferior good (Koutsoyiannis, 1987). Assuming health is a normal good, those with higher incomes were hypothesised to give higher values.

The above hypotheses were tested by comparing means of different categories within each factor as well as differences in the distribution of values to determine if VAS and TTO values for each disease state varied between groups. The Mann-Whitney and Kruskal Wallis non-parametric tests were used to test the hypotheses. In addition, the *t-test statistic* was used to test for differences between means for variables with two categories and the one-way ANOVA *F-statistic* for variables with more than two categories, for comparison. To test for equality of matched pairs of observations (i.e. pairs of VAS-TTO values) the Wilcoxon matched pairs signed rank test was used, which test the null hypothesis that both distributions are the same (StataCorp.2001, p.213).

8.1.5.3 Reliability

Reliability was assessed using the test retest and inter-rater reliability for the VAS and TTO values for disease states. Inter-rater reliability requires comparisons of ratings for the same subject from different interviewers on different occasions. These ideal conditions were not met for this study and inter-rater reliability was assessed by comparing mean values for disease states obtained from groups of interviewees by interviewers during the initial and retest valuations separately. The Kruskal Wallis test and one way analysis of variance F test were used to test for significant differences in mean values between interviewers.

To assess test retest reliability, responses on some selected variables from the same respondent at initial and retest valuation were compared. Spearman's correlation coefficients were obtained for health related variables together with comparison of values between test and retest. Variables related to practicality in use (such as number of times task explained, ease of use and comparison of techniques, extent of thinking and preferred technique) of the valuation techniques were also correlated to test whether the respondents were consistent in their responses between test and re-test.

Comparing values for the disease states obtained two weeks apart was used to assess test-retest reliability. Tests of differences in mean values between test and retest were done using the paired *t-test* statistics and the Wilcoxon test for paired groups. Correlations between test and retest values were also done using the Spearman's correlation coefficient. Correlations were expected to be above 0.5 to indicate reliability (Bowling, 1997). Intra-rater reliability was undertaken for both techniques using the Kappa coefficient of agreement, which lies between 1 (perfect agreement) and 0 (agreement by chance). For intermediate values Landis and Koch (1977a, 165) in Statacorp. (2001) suggested the following interpretation: below 0.0 - poor; 0.00-0.20 - slight; 0.21-0.40- fair; 0.41-0.6-moderate; 0.61-0.8 -substantial; and 0.81-1 - almost perfect.

8. 2 RESULTS

8.2.1 Characteristics of Study Subjects

8.2.1.1 Socio-economic and demographic characteristics

Table 8.3 presents socio-economic and demographic characteristics of community and patient samples. Although the age range and standard deviations were not very different between the community sample and the patients, the mean age amongst community members was significantly higher ($p<0.01$) than that of patients. The respondents in both samples were predominantly female, married, and the majority had children. There were significantly ($p<0.05$) more respondents with children amongst community members compared to patients, although the mean number of children was not significantly different.

While there were more without any education amongst the community subjects, there were also more with above secondary level education whilst majority of patients had primary level education with fewer reporting none. The majority were farmers in both samples although the patient sample had over 15% students and about 10% business persons and casual labourers each. The Mann-Whitney test indicate that the distribution of occupation amongst community and patient respondents was significantly different at $p<0.05$.

Monthly total household expenditure was assessed by asking about monthly expenditure on individual items which included food, clothing, rent, school fees, health, debt repayment, fuel, leisure, farming and any other expenditure. This was in recognition of the unreliability of asking about monthly income directly and people's reluctance to divulge such information. However, respondents were also asked to indicate against pre-determined income groups, the one that best described their monthly income. The majority (68.3% to 71.93%) of both patients and community members reported monthly incomes below Kshs 10,000. Patients reported significantly lower monthly expenditures than community members ($p<0.01$).

8.2.1.2 HRQL and illness status of study subjects

Amongst six HRQL domains specified on 5 levels from 'none of the time' through 'a little, some, most' and 'all the time', subjects were asked to indicate the level that best described their HRQL in the last two weeks. This information was used to compute a HRQL index by adding up the levels and classifying HRQL states as: 1-6, perfect health, 7-12, mild state, 13-18 moderate state, 19-24, severe state and 25-30 as very severe state. Respondents were also asked to rate on the VAS scale, their health state in the last two weeks as well as their usual health state.

The majority of community members (above 63%) and patients (above 53%) indicated that none of the HRQL domains was affected any of the time (table A8.1 in appendix 8.4). Between 10% and 23% of community members and patients reported all HRQL domains except social participation as being affected some of the time. Seven to 10% of community members reported mobility, performance of output and work and feeling of strength and energy being affected most or all of the time. Similarly, 7% to 15% of patients reported performance of daily duties, performance of output and work and feeling of strength and energy being affected most or all of the time. Tests[66] for significant differences between the distribution of patients and community members across frequency levels revealed no differences, suggesting that the two groups were from a population with the same distribution of how HRQL domains were affected. Figure A8.1 (appendix 8.4) shows that the proportion of community and patient members who reported perfect, mild, moderate and severe HRQL states were nearly similar with majority (49% and 47% respectively) reporting perfect health.

Table 8.4 presents mean VAS ratings of usual health and health state in the last two weeks. In general the mean VAS ratings are high, indicating good health, in all cases. VAS ratings of health state in the last two weeks among community members were statistically significantly higher ($p<0.05$) than those of patients. There was no evidence of differences in

[66] These were the Mann-Whitney test, Spearman's rho and Pearson's chi^2.

Table 8.3: Socio-economic and demographic characteristics for community and patient valuation samples (%)

	Patient sample (n=60)	Community sample (n=57)
Age: (years) Mean (SD) [range]	32.7 (15.9) [15-75]*	41.4 (17) [17-77]*
Gender (%)		
Males	41.67	38.60
Females	58.33	61.40
Marital status (%)		
Single	30.00	12.28
Married	58.33	77.19
Separated / divorced	5.00	5.26
Widow / widower	6.67	5.26
No. of children		
% with children	70.00*	87.72*
Mean (SD) [range]	3.6 (2.7) [1-10]	4.6 (3.3) [1-15]
Education level (%)		
None	5.00	21.05
Primary	78.33	47.37
Secondary (O and A level)	16.67	25.31
Degree	0.00	1.75
Adult education	0.00	3.51
Occupation (%)		
Farmer	61.67*	84.21*
Teacher	0.00	1.75
Business person	10.00	5.26
Casual labourer	11.67	3.51
Civil servant	0.00	0.00
Student	15.00	3.51
Other	1.67	1.75
INCOME (KSHS): (%)		
0-2,000	23.33	21.05
2,001-5,000	36.67	28.07
5,001-10,000	8.33	22.811
10,001-20,000	3.33	14.04
20,001-50,000	3.33	3.51
Above 50,000	1.67	0.00
No response	1.67	1.75
Not applicable	21.67	8.77
MONTHLY EXPENDITURE (KSHS): MEAN (S.D.) [RANGE]		
Total expenditure `000	5.8 (5.12) [0-19.7]*	11.6 (9.1) [0-37.7]*

Tests for significant differences used the Mann-Whitney test * significant at $p<0.05$.

VAS ratings of usual health between the two groups. According to these reports the health status for the community group was higher than for the patients and was not different from the usual in the last two weeks. However, patients' health status in the last two weeks was significantly ($p<0.01$) lower than their usual health status. These results seem to suggest that the patient group perceived their health status to be lower.

Table 8.4: VAS ratings of usual health state and health state in the last two weeks (mean (SD) [range]) [Mann-Whitney test]

	Patients (n=60)	Community (n=57)
Health state in the last two weeks	73.1 (21.2) [20-100]	78.5 (24.9) [4-100]
Usual health state	85.8 (13.6) [50-100]	83.2 (21.2) [2-100]

Table 8.5 shows the proportion of respondents reporting illness in the last two weeks prior to the interview as well as during the interview. The Fisher's exact chi^2 statistic shows that the proportion of patients and community members reporting illness during the interviews were not significantly different, while significantly more patients reported being ill in the last two weeks. This result corresponds to the finding that more of patients had lower health status in the last two weeks. However, it appears that during the interviews, both samples were similar in their reports of illness.

Table 8.5: Illness status currently and in last two weeks

	Patients (n=60)	Community(n=57)
% currently ill	36.67	22.81
% aware what current illness was	28.33	22.81
% ill in last two week s	56.67*	33.33
% aware what illness in last two weeks was	46.67*	31.58

Fisher's exact chi². * P<0.05

8.2.2 Practicality

Table 8.6 shows various forms through which practicality for VAS and TTO was assessed. For each aspect, four types of comparisons were made; two within and two across the groups. The first two comparisons assessed whether there were any statistically significant differences between proportions for the VAS and TTO within patients and community, using the Wilcoxon sign rank test for paired groups (i.e. comparing column A and B; and C and D). The third and fourth comparisons tested for differences in proportions for the VAS and TTO between patient and community samples, using the Mann-Whitney test for unpaired groups (i.e. comparing columns A and C; and B and D). The comparisons are based on n=60 patients and n=57 community members, rather than the pooled data set.

For the majority of subjects, the VAS and TTO only had to be explained once. VAS required significantly more explaining than the TTO ($p<0.05$) for community members. Community members also required significantly more ($p<0.01$) explaining in using VAS than patients. The majority of patients and community members rated VAS as easy to fairly easy to use with about 20% of patients rating it as difficult. Similarly TTO was rated easy to fairly easy to use by fewer community members compared to patients. 36.8% of community members rated TTO as difficult. In general, significantly more community members rated TTO as difficult compared to VAS ($p<0.01$). The distribution of ratings of ease of understanding VAS and TTO were not different amongst patients.

Respondents were asked to state difficulties experienced during use of the VAS and TTO. The majority of patients and community members experienced no difficulties using VAS. However, more than half of community members and patients experienced difficulties using TTO. There was strong evidence suggesting that more patients ($p<0.01$) and community members ($p<0.05$) found the TTO difficult than VAS. The proportions finding the VAS difficult amongst the patients and community members were not significantly different. Difficulties associated with VAS included inability to differentiate between disease state scenarios and difficulties in remembering what was in a previous scenario as

well as difficulties in comparing and ranking the disease states. The most commonly cited difficulties associated with TTO were comparing and choosing between the two alternative lives and also comparing time in both lives. Other less cited difficulties included: comparing and differentiating the disease states; visualising and imagining people in these disease states; required a great deal of thinking and difficulties in choosing due to not knowing what will happen after the duration specified.

The majority of patients and community members (51% and 58%) rated VAS easy compared to below 37% who rated TTO easy. A slightly higher proportion (>20%) of community members rated TTO as difficult. Statistical evidence suggests that VAS was rated easy by most of the patients ($p<0.05$) and community members ($p<0.01$), indicating that VAS was considered easier than TTO. Asked to indicate their preferred technique, the majority of community members preferred the VAS while more than half of the patients preferred the TTO. The chi square test of association between choice of a technique and age, gender, marital status and education level did not reveal any association, indicating that none of these groupings of people preferred a given technique.

Ease of understanding and use of the techniques was also assessed by the interviewers. The majority of both patients and community members were rated as having no difficulties although the proportions are lower for TTO and more so amongst community members. Consequently, more community members (42%) were rated as having little to somewhat difficulty in understanding TTO compared to patients (33.4%). Neither the Mann-Whitney nor Wilcoxon tests indicated statistically significant differences in the distribution of respondents' level of difficulty in understanding the VAS and TTO, when compared across or within patient and community groups.

Table 8.6: Ease of understanding and use of VAS and TTO (%)

		Patients(n=60)		Community(n=57)	
		VAS	TTO	VAS	TTO
	Column:	A	B	C	D
No. of times task explained					
	1	96.7	96.7	82.5	91.2
	2	1.7	3.3	12.3	8.8
	3	1.7	0.0	3.5	0.0
	4	0.0	0.0	1.8	0.0
Ease of understanding					
	Very easy	1.7	0.0	7.0	5.3
	Easy	50.0	45.0	54.4	35.1
	Fairly easy	26.7	31.7	26.3	22.8
	Difficult	20.0	21.7	10.5	36.8
	Very difficult	1.7	1.7	1.8	0.0
% experiencing difficulties					
	None	63.3	50.0	68.4	43.9
	Others	36.7	50.0	31.6	56.1
Interviewer's rating of respondent's level of difficult in understanding task					
	None	75.0	65.0	66.7	54.4
	A little	21.7	26.7	15.8	22.8
	Somewhat	1.6	6.7	8.7	19.3
	Difficult	0.0	1.7	5.3	3.5
	Very	1.7	0.0	3.5	0.0
Interviewer's rating of respondent's level of comfort with the task					
	None	0.0	0.0	3.5	0.0
	A little	1.7	3.3	3.5	1.8
	Somewhat	5.0	10.0	7.0	21.1
	Comfortable	33.3	41.7	35.1	29.8
	Very	60.0	45.0	50.9	47.4
Interviewer's impression of interview and responses obtained					
	Very good	58.3**	30.0	59.7**	28.1
	Good	35.0	40.0	31.6	38.6
	Satisfactory	5.0	25.0	5.3	28.1
	Poor	0.0	5.0	0.0	5.3
	Very poor	1.7	0.0	3.5	0.0
Comparisons of VAS and TTO by respondents					
	Easy	51.7	36.7	57.9	35.1
	Neither easy nor difficult	41.7	48.3	31.6	40.35
	Difficult	6.7	15.0	10.5	24.6
Preferred technique by respondents		46.7	53.3	71.9	28.1

** Significant at $p<0.01$ using Wilcoxon sign rank test

The majority of patients and community members were rated by the interviewers as being comfortable and very comfortable during the valuation task, with 21% being rated as somewhat comfortable in using TTO. There were no differences in the distribution of respondents' level of comfort amongst patients and community members in using the VAS and TTO. 93% and 91% of VAS responses were rated by the interviewers as good to very good for patients and community members respectively while for the TTO these were 70% and 67%. Over a quarter of both VAS and TTO responses were considered satisfactory by the interviewers. In general, less than 5.3% of subjects were considered to provide poor to very poor responses. There was strong statistical evidence that the proportion of community members ($p<0.01$) and patients ($p<0.01$) providing good to very good responses was higher for the VAS than the TTO.

To understand the types of people who found the VAS and TTO less practical, an examination of the above aspects of practicality was done. Respondents who required the use of technique to be explained over 3 times, said the task was very difficult; whom the interviewer judged as finding the task difficult and very difficult, who had none or little comfort in using the techniques and whose responses the interviewer judged as poor to very poor were categorised by gender, age, marital status and education level. Findings showed that amongst the patients, only one young (31-45 years) female with primary education experienced problems with VAS while 8.3 % of all the respondents had problems with TTO. 60% of these patients were older (over 46 years), married males with primary education. Amongst community members 14% and 11% had problems with VAS and TTO respectively. For the VAS, these people tended to be older (over 46 years), married males with primary or no education, while for the TTO they tended to be married females of all ages with no education or primary education only.

Table A8.2 in appendix 8.4 presents other aspects of practicality that were assessed. In terms of extent of thinking required, the majority of both patients and community members only needed to think a little to very little to complete the VAS but needed to think hard to very hard to complete the TTO. More community members ($p<0.05$) required less effort to

use the VAS than TTO. This pattern was not statistically significant amongst patients ($p<0.1$).

The mean time to value six states using VAS was 9 minutes for the community members compared to 6.9 minutes for patients. Similarly for the TTO, community members took statistically significantly more time than patients to value four disease states (VAS ($p<0.01$) and TTO ($p<0.01$)). There was also strong evidence amongst both patients ($p<0.01$) and community members ($p<0.01$) that the TTO took more time than VAS.

Respondents were asked to state the types of people who could not use these techniques using an open-ended question. Over 90% mentioned children and over 75% the very old people. Other types of people mentioned included the illiterate, those who are sick, the mentally unstable and drug abusers. The mean age of children who could not be asked to use the valuation techniques was given as approximately 9-10 years. The maximum age was judged to be above 51 years by community members and above 63 years by patients.

8.2.3 Construct Validity of techniques

8.2.3.1 Distribution of VAS and TTO values for disease states

Table A8.3 in appendix 8.4 presents means and measures of dispersion for VAS and TTO values for disease states while figure A8.2 in appendix 8.4 compares those distributions for TTO and VAS values by disease state. The mean VAS values for all disease states were positive and decreasing as states worsened. For all the disease states, standard deviations were small such that the 95% confidence intervals were narrow, implying that values were not very widely spread for VAS. The range of values indicated that there were cases where disease states were assigned extreme values. The mean VAS value for the state perfect health was 0.96, although the majority (73.5%) assigned perfect health a value of 1.

The mean TTO values for the mild, moderate and severe disease states were positive and decreasing as the states worsened with the very severe disease state being assigned a value

of 0. There was more variability in the TTO values for the disease states, resulting in very large confidence intervals. Figure A8.2 in appendix 8.4 shows that the VAS values were less spread and tended towards a normal distribution compared with TTO values which tended to span the entire scale with variability increasing as disease state severity worsened. For the mild and moderate disease states the TTO values were higher than the VAS values while the reverse held for the severe and very severe disease states.

8.2.3.2 Ranking of states by VAS and TTO

Ranking of disease states according to mean VAS and TTO values followed the logically expected ranking where the mild state was valued better than the moderate and likewise for severe and very severe states. Hence, worse off disease states were assigned lower values. Although the mean values suggest that the disease states were ranked logically, there were a few individuals with inconsistent rankings of disease severity. Table A8.4 (appendix 8.4) presents information of inconsistencies in ranking disease states. The majority (92%) rated all the six VAS scenarios logically, with the mean number of states ranked according to severity being six. There was more inconsistency in ranking using TTO, where only a fifth of the respondents ranked all the four disease states according to severity and 27% ranked either one or none of the states according to the expected logic. On average, about two states were ranked according to disease severity using the TTO. An examination of rankings with the VAS and TTO revealed that the majority (91%) of the 27% who ranked either one or none of the states logically had ranked all the six states using the VAS logically. The Wilcoxon sign rank test showed that the distribution of inconsistencies using the VAS was significantly ($p<0.01$) different from that of the TTO. The 27% who ranked none or one state according to disease severity using the TTO were investigated. The majority were married (59%) females (66%), aged less than 45 years (72%) with no education or primary education only (85%).

The very severe disease state (F) was the most inconsistently ranked, but more so using the TTO. This situation was to a lesser extent reflected for the severe state (E). Those who considered the mild and moderate state worse than death said that the problems in those

states were too many to persevere. For the severe state and very severe states, the majority (9.4% and 21%) considered the problems too many to persevere and that such a state would lead to total dependency on others (1.7% and 4.3%). Four point three percent and 6.8% considered that being in the severe and very severe disease states respectively, would make them unproductive and would be doing nothing in this world, hence better off dead.

8.2.3.3 Differences between mean VAS and TTO values

Given the similarity of disease states being valued, it was expected that values from both techniques for a given state would be similar. Table 8.7 presents means and differences between means of VAS and TTO values. Tests for differences between mean values using *t-tests* and Wilcoxon sign rank tests shows that the TTO values are significantly higher for the mild and moderate disease states and significantly lower for the very severe disease state. The difference in means for the severe state was only significant using the *t-test* but not the Wilcoxon test.

Table 8.7: Construct validity: Tests of equality of mean VAS and TTO values (n=117)

STATES	Mean VAS	Mean TTO	Mean difference (VAS-TTO)
State A	0.67	0.73	-0.07[ab]
State C	0.48	0.55	-0.07[ab]
State E	0.32	0.24	0.08[a]
State F	0.18	0.00	0.18[ab]

a- $p<0.05$ using t test. b – $p<0.05$ using Wilcoxon sign rank test

8.2.3.4 Pearson's correlation between VAS and TTO values

Table 8.8 presents Spearman's correlation coefficients between VAS and TTO values for disease states. Positive correlations were expected, signifying construct validity if VAS and TTO measured the same construct. There were mixed findings with negative, very low and insignificant correlations for the mild and moderate disease states and positive, low and insignificant correlations for the severe and very severe states. These findings suggest that

for the severe and very severe states, the VAS and TTO values seem to represent the same construct as opposed to the mild and moderate states.

Table 8.8: Spearman's correlation coefficient between VAS and TTO values for disease states (n=117)

STATES	Spearman's
State A	-0.10
State C	-0.05
State E	0.14
State F	0.11

No significant correlations at 0.05 level.

8.2.3.5 Relationships between VAS and TTO values and rater characteristics

Tables A8.5 to A8.16 in appendix 8.4, present results of tests of hypotheses on how VAS and TTO values varied by characteristics of respondents. These included; gender, age, marital status, whether rater has children or not, number of children, education level, monthly income and expenditure, whether ill or not during valuation, experience with schistosomiasis, VAS rating of own health status in the last two weeks and usual health status. Only the non-parametric tests (Mann-Whitney and Kruskal Wallis) are reported in this section, although tables A8.5 to A8.16 include the *t* test and the one-way ANOVA *F* test for comparisons.

The results indicate that VAS and TTO values for all disease states did not vary by age, marital status, whether one had children or not and the number of children, monthly expenditure and income or VAS rating of usual health. The VAS values for the severe and very severe disease states and TTO values for the mild and severe states also did not vary with any of the characteristics tested for. The VAS values for the mild disease state were significantly ($p<0.05$) higher for females, those with no education, and whose VAS ratings of health status in the last two weeks was higher. The VAS values for the moderate disease state were significantly ($p<0.05$) higher for those with primary or no education and those whose VAS ratings of health status in the last two weeks was lower. The TTO values for the moderate disease state were significantly ($p<0.05$) lower for those who were ill during

the valuation and those who had had schistosomiasis mansoni a year ago. Those who had rated their health status in the last two weeks very low assigned a significantly ($p<0.05$) high value to the very severe disease state suggesting that they considered their health status much worse than the severe state. In general the factors that appeared to have an effect on values were gender, level of education, presence of illness, experience with the disease and people's perception and valuation of their own health status.

8.2.4 Reliability of VAS and TTO valuation techniques

8.2.4.1 Inter-rater reliability

As noted earlier, assessment of inter-rater reliability was done by comparing the mean VAS and TTO values for disease states from different respondents by different interviewers. Table 8.9 shows the mean VAS and TTO values at initial and retest valuation by four interviewers (raters). On both occasions, the Kruskal Wallis and one way analysis of variance tests showed no statistically significant differences in values for disease states at both initial and retest valuation, indicating stability of values for same disease states across interviewers.

Table 8.9: Inter-rater Reliability: Mean values for VAS and TTO values at initial and retest valuation by rater

	INITIAL VALUATION (n=117)							
	VAS				TTO			
	STATE A	STATE C	STATE E	STATE F	STATE A	STATE C	STATE E	STATE F
RATER 1	0.70	0.51	0.33	0.17	0.70	0.56	0.38	0.12
RATER 2	0.68	0.47	0.32	0.19	0.80	0.54	0.26	0.01
RATER 3	0.65	0.46	0.31	0.17	0.72	0.57	0.11	-0.12
RATER 4	0.65	0.48	0.34	0.19	0.74	0.55	0.25	-0.00
	RETEST VALUATION (n=60)							
	VAS				TTO			
	STATE A	STATE C	STATE E	STATE F	STATE A	STATE C	STATE E	STATE F
RATER 1	0.72	0.51	0.35	0.20	0.78	0.72	0.25	0.17
RATER 2	0.66	0.45	0.32	0.22	0.81	0.64	0.43	0.26
RATER 3	0.68	0.51	0.30	0.21	0.76	0.49	0.38	0.23
RATER 4	0.63	0.48	0.34	0.21	0.77	0.61	0.37	0.27

No significant differences using one-way ANOVA F statistic and Kruskal Wallis test.

8.2.4.2 Test-retest reliability

The re-test sample for assessment of reliability comprised of 60 respondents. Their mean age was 35.7 years. Table A8.17 in appendix 8.4 shows that the majority were females, married, and had children (mean: 4 children). The majority had primary and secondary education, were mainly farmers, with a mean monthly expenditure of 9,600 Kenya shillings. The Mann Whitney test showed that the test retest sample (n=60) was not statistically significantly different (p values for all variables greater than 0.09) from the initial valuation sample (n=117) from which it was sampled in terms of socio-economic and demographic variables.

Table A8.18 in appendix 8.4 shows tests of differences in mean ratings of usual health status as well as health status in the last two weeks prior to the valuation between test and retest were not significantly different. Similar findings were found with respect to whether one had illness two weeks prior to the interview as well as during the interview. The correlation coefficients for the usual health status and health status in the last two weeks were 0.65 and 0.52 respectively and were statistically significant at $p<0.01$ level. Although the correlation for presence of illness in the last two weeks and during the interview were low (0.37), it was significant at $p<0.01$ level. This finding shows that in the two weeks between test and retest, there wasn't a significant change in the distribution of reports of health status or illness. However, the moderate correlations for reports of illness imply that those who were ill during the initial interview were not necessarily the same as those who reported illness at retest.

Other variables that were investigated to shed light on test retest reliability related to use of the valuation techniques. Table A8.19 in appendix 8.4 shows that there was a significant reduction in the number of times the VAS procedure was explained at initial and retest valuation. However, there were no significant differences in the distribution of ease of using VAS, its ease compared to TTO, the extent of thinking required to complete the VAS task or preference of VAS compared to TTO. There were also no significant differences in the distribution of these variables with respect to TTO. However, the Spearman's

correlation coefficients were low for the TTO, although the coefficient on ease of TTO compared to VAS was significant. For the VAS the coefficients on the number of times the procedure was explained, ease of using the technique and extent of thinking required can only be considered moderate although significant.

Table A8.20 in appendix 8.4 shows the distribution of VAS and TTO values between test and re-test. Both the initial and retest VAS values were positive and decreasing with increasing disease severity, whereas only the TTO retest values followed this pattern. Initial TTO values for the mild, moderate and severe states were positive while the value for the very severe state was negative. However, the values decreased with increasing severity of the disease states. The variability in the VAS values was low with a narrow 95% confidence interval. However, variability in TTO values was large and tending to increase with increasing disease severity. The 95% confidence intervals are broad.

Table 8.10 below presents results of test-retest reliability of VAS and TTO for differences between mean values. The VAS values for all disease states except the severe state were slightly higher (mean difference less than 0.03) at retest. The TTO values for the severe and very severe disease states were higher and for the mild and moderate disease states lower at retest. The findings revealed that the VAS values were reliable for the mild and severe disease state as were the TTO values for all states except the very severe disease state. Both the VAS and TTO values for the very severe disease state were significantly higher at retest than at initial valuation, indicating unreliability in these values. VAS values for the moderate state were also significantly higher at retest.

Table 8.10: Mean differences between initial and retest VAS and TTO values for disease states (n=60)

	VAS			TTO		
	Initial mean	Retest mean	Difference in means	Initial mean	Retest mean	Difference in means
State A	0.64	0.67	0.03	0.80	0.78	-0.02
State C	0.46	0.49	0.03[b]	0.63	0.61	-0.02
State E	0.33	0.32	-0.01	0.29	0.35	0.06
State F	0.19	0.21	0.02[ab]	-0.02	0.23	0.25[ab]

a- $p<0.05$ using t test. b – $p<0.05$ using Wilcoxon sign rank test

Table 8.11 shows that the Spearman's correlation coefficients between initial and retest VAS values for all disease states were above 0.57. The null hypotheses that the test and retest values were statistically different was rejected at $p<0.001$, suggesting reliability for the VAS values. The correlation coefficient for the TTO values for the mild and very severe states were above 0.40 and significant, also suggesting reliability. However, the null hypotheses that test and retest values for the moderate and severe disease states were statistically different was not rejected ($p>0.1$) suggesting unreliability for these TTO values.

Table 8.11: Spearman's correlation and Kappa coefficient of agreement between initial and retest VAS and TTO values for disease states (n=60)

	VAS		TTO	
	Spearman's correlation	Kappa coefficient	Spearman's correlation	Kappa coefficient
STATE A	0.59[a]	0.20	0.40[a]	0.15
STATE C	0.59[a]	0.15	0.14	0.12
STATE E	0.57[a]	0.21	0.20	0.09
STATE F	0.59[a]	0.30	0.46[a]	0.07

[a] $p<0.001$ All Kappa coefficients significant at $p<0.005$ except TTO coefficient for state F.

The Kappa coefficient for agreement between initial and retest VAS values for all disease states was fair (Statacorp.2001) and statistically significant at $p<0.005$ (table 8.11). In the initial and retest TTO values, there was slight agreement in values for all the disease states. The kappa coefficient was significant at $p<0.001$ for all the disease states except the very severe one, which was significant at $p<0.05$. Hence, the kappa coefficient suggests that both the VAS and TTO values were reliable, despite the size of the coefficient of agreement.

8. 3 DISCUSSION

The discussion addresses how practical, construct valid and reliable the VAS and TTO were when used in Kenya and what might have influenced these aspects in this setting. First, the practicality of the instruments is discussed followed by construct validity. To explore construct validity, a discussion on evidence of ordinal rankings and tests of hypothesis is presented together with levels of consistency achieved by valuation techniques. In addition, a discussion on variation in values is also presented as further evidence on construct validity of the tools. Finally, the discussion on reliability is presented. Exploration of these issues sheds light on the extent to which these tools could be reasonably used for valuation of disease and health states in developing countries for use in health care resource allocation and decision making.

Assessment of practicality in this study adopted a slightly different approach to that used in other studies as described in section 8.1.5.1. For example Brazier et al. (1999b), reviewed practicality in terms of acceptability (length, complexity and respondent interest), response rates (indicating level of agreement to participate) and completeness (indicated by level of missing data). Others have used comprehension rates (Tan Torres, 1991), completion time (Mohide et al. 1988; Buxton and Ashby, 1988; Unic et al. 1998; Stigglebout et al. 1994). While most of the above aspects are observable, this study engaged the respondents deeper in a detailed assessment of more aspects. These aspects included the number of times the task was explained, eliciting opinions on different aspects of appropriateness and types of people who cannot use the tools, in addition to completion time, ease of use and extent of thinking effort required in completing the task among others.

A comparison of performance of the VAS and TTO on aspects of practicality reveals that the VAS was better than TTO on all aspects except the number of times required to explain the task before the respondent was deemed to understand it. The VAS was considered easy by the majority, although the patient group seemed to consider both techniques similar in terms of ease of understanding. In terms of the preferred technique, the community group choose the VAS while more patients choose the TTO. Respondents experienced difficulties

with both techniques, but there were more difficulties related to the TTO. The main difficulties related to retention of information in scenarios, comparing them to assign them values as well as choosing between them. If all the information in a scenario is not processed and used in assigning them values, then it remains questionable whether such values are appropriate for incorporation into decision making. It also raises questions about "what is it that respondents take into account in assigning a value to a disease/ health state?" Perhaps further research should attempt to unravel, through reflection and perhaps con-joint analysis, what pieces of information in a given scenario influence its value.

That the VAS required more explaining is perhaps a reflection of the fact that the order in which the techniques were used was not randomised. The VAS was presented first followed by the TTO in all the interviews, which might imply that by the time the TTO was introduced, the respondents already had the gist of the task and therefore needed less explaining. Had the order been randomised, perhaps the number of times the technique was explained would have been different, but this could be the subject of future studies.

In terms of other aspects like interviewer's judgement on ease of understanding the technique, level of comfort in undergoing the task, quality of responses, extent of thinking required and time to complete the valuation task, the VAS performed better than the TTO. Given these findings, which technique can be considered practical for use in Kenya? It would appear that VAS would be a more practical technique based on its ease of use, as it takes less time, is voted for by the majority, is less mentally demanding and fewer respondents experience difficulties using it. This finding is consistent with the literature (Bleichrodt et al. 1997; Dolan and Sutton, 1997; Gold et al. 1996; Brazier et al. 1999b), that the VAS performs marginally better than other techniques and also has cost advantages. However, the results also suggest that the TTO would also be a practical technique to use in Kenya, considering that the majority understood its use quickly, found it easy to fairly easy, were comfortable in using it and was the preferred technique by the patient group. It could also be argued that, this being the first time ever to use this technique in Kenya, its repeated use could improve its level of practicality considerably. It is therefore suggested that the technique be tested further with a view for future use in valuation of health outcomes in this

setting. It was suggested that the very old and children below 10 years would not be able to use these techniques. While these techniques were not tested on those below 15 years, further research should look into feasibility of use of these techniques with younger age groups in this setting. The findings in this thesis therefore challenge previous assertions (Kirigia, 1994) that the TTO cannot be used amongst populations with relatively low education.

The evidence adduced from assessing construct validity of both the VAS and the TTO was somewhat mixed. Ranking of disease states using mean values for groups as a whole followed a logical sequence where the worse off disease states were assigned lower values using both techniques. Hence, both the VAS and TTO were able to predict the ranking of the disease states according to their severity providing evidence of their ordinal properties (Brazier et al. 1999b). This finding suggests that the valuation given to a disease state is a decreasing function of its severity (Dolan, 1996) and supports construct validity for both the TTO and the VAS. However, the distribution of values for the VAS and the TTO was different. VAS values tended to have a normal distribution, with low, nearly constant standard deviations and narrow confidence intervals as opposed to the TTO values whose confidence intervals were wide, with variability increasing with severity of disease state. This implies that respondents held widely differing opinions of severity of the disease states with more variation as disease severity worsened for the TTO.

An examination of differences in values for immediately following disease states (i.e. A-C = 0.19, C-E = 0.16 and E-F =0.14) showed that VAS states were almost placed equidistance from each other, suggesting use of the entire scale. Response spreading challenges the VAS in terms of its interval scale properties, and numbers produced using the scale may not be meaningful (Read et al. 1984; Bleichrodt and Johannesson, 1997). However, response spreading can be controlled for by valuing one state at a time (Kaplan et al. 1993) and stressing to the subject that relative difference between states is an issue, to capture the interval scale properties (Torrance, 1986). The TTO values did not possess this characteristic (A-C = 0.18, C-E = 0.31 and E-F =0.24), perhaps due to the fact that each state is valued separately and a reflection of interval scale properties of the tool. These

features together with the wide range of values for each disease state may cast some doubt about the validity of these instruments to provide valid values for the disease states.

A notable point was that the mean VAS value for the state described as perfect health was well below 1.0 as a quarter of the respondents assigned it a value less than 1.0. This confirmed opinions[67] that the concept of perfect health was not easily conceptualised and comprehensible to some and might mean that the endpoint of the VAS scale means something different to different people. This touches on issues of validity of the scale especially when the assumption is made that 1 represents the best imaginable state (usually perfect health as in the case of TTO). This underscores the importance of certifying the extent of content validity and conceptual equivalence of the scale across cultures. A clear understanding of the conceptualisation of the anchor states in different settings is needed to determine whether values require re-scaling or not, while providing the ingredients of the anchor states (i.e. perfect health and death) that would be easily related to by people in a given cultural milieu.

Although the mean VAS and TTO values provided a logical ranking consistent with disease severity, detailed examination revealed that there were some logical inconsistencies. This was a common problem with the TTO where only a fifth of respondents provided logical rankings for all the four disease states. The VAS had the lowest level of logical inconsistency compared to the TTO, a finding consistent with Mahapatra et al. (2002). Dolan and Kind (1996) distinguish between primary (arising from limitations of the respondents) and secondary (due to measurement procedure) inconsistencies. Secondary inconsistencies arising from preference reversals have been termed as 'consistent' and that they do not necessarily suggest that the respondent is non-transitive in their choice (Loomes and Sugden, 1982; Dolan and Kind 1996). Primary inconsistencies may however arise due to confusion and poor understanding of the valuation task by the respondent, and may imply not only that the technique is impractical but that it is also invalid.

[67] These observations were addressed in chapter 7.

That those who had logical inconsistencies using the TTO had no inconsistencies using the VAS may suggest that the inconsistencies were due to difficulties and or confusion in using the technique rather than the disease states being perceived and therefore valued differently, which in itself casts doubt on the validity of the TTO. Considering the proportions of respondents citing difficulties with the use of TTO and looking at the range of values for same disease states seems to suggest that the validity of the TTO may have been compromised due to poor understanding and or confusion with the actual use of technique. It could also be that the differences in the level of inconsistency between VAS and TTO were as a result of the differences in certain aspects embodied in the techniques that respondents considered while using the two techniques. For instance, the trading of time (quantity) for health (quality) may have encouraged respondents to reflect more on their choices by considering what they would achieve by being in perfect health as opposed to being in the given disease state, resulting in quite different values for a similar state. It would therefore be pre-mature at this point to declare the TTO invalid in favour of the VAS, without further investigation of the causes of these apparent inconsistencies and further investigation of other factors likely to cause differences. If deemed invalid due to lack of understanding and confusion, it implies that perhaps with more use the validity would improve as respondents become more familiar with the technique. Nevertheless, the VAS was both practical and provided logical ranking of disease states when used in this setting, albeit with the suspicion of response spreading.

Assessment of construct validity via comparison of means and correlation analysis provided little evidence for either the VAS or the TTO. Differences in mean TTO and VAS values were statistically significant. TTO values for the mild and moderate states were higher than VAS values and vice versa for the severe and very severe states. That TTO values for the severe and very severe states are much lower may suggest that the respondents were willing to trade more time to return to better health while for the mild and moderate states, they were less inclined to trade more time. While differences in mean values lend no support for construct validity, it can also be argued that this might not be the best way to assess validity of valuation techniques given that previous studies have

documented that values differ by method (Robinson et al. 1997; Read et al. 1984; Krabbe et al. 1997; Torrance et al.2001; Dolan and Sutton, 1997; Froberg and Kane 1989b).

Positive but low and insignificant correlations (0.14 and 0.11) were obtained for the severe and very severe states respectively. These values are comparable with Buckingham et al's. (1996) values for construct validity for TTO that ranged between 0.12-0.22. Construct validity for the mild and moderate states using correlation analysis was not supported. These findings could be attributed to several factors. Construct validation through correlation analysis assesses both theory and the measure at the same time (Streiner and Norman, 1995). Hence low correlations could result from incorrect theoretical context and hence wrong specification of constructs, or use of incorrect statistical technique, or errors in measurement, or that the measure lacks validity for the particular theoretical construct (Carmines and Zeller, 1979; Streiner and Norman, 1995).

Considering that the two techniques ordered the discase states according to their severity could be used to exonerate theory as the culprit. Also, despite the inconsistency reported at the individual level, at the group level there seemed to be evidence of consistency (transitivity) in choosing, a further justification of theoretical constructs. Since construct validation is an on-going process, it would be pre-mature to declare the VAS and TTO invalid at this stage of their use in Kenya. Two possible explanations for the poor correlations are errors in measurement and the statistical technique used.

Measurement errors can arise from both the respondents and interviewers in using the tool. Although interviewers were trained extensively on how to administer the instruments, errors cannot be ruled out. Also the respondents might have had difficulties with the use of the tools especially as evidenced with the TTO, thereby resulting in errors in values they attached to disease states. The phenomenon of valuing disease states was new in this area and this might have biased the use of the instruments. Although this effect is unknown, more use and familiarisation with the concept of valuing disease/health together with the instruments used, might improve their validity. Another possibility that could have introduced measurement error was that respondents attached very different values to the

disease states because they valued different aspects of the disease state scenarios i.e. symptoms and HRQL domains differently. Hence, the measurement bias may have arisen from respondents' perceptions and valuation of different aspects of disease state scenarios, rather than lack of validity of the tools.

Lastly, the low correlations could be due to the statistical technique used. Although correlation analysis is the commonest tool for testing validity, it might be the case that the two techniques were capturing different aspects of the construct. The VAS has been found to correlate poorly with choice based methods such as TTO and SG (Zug et al. 1995; Rutten van Molken et al. 1995; Brazier et al. 1999b) and correlate strongly with measures of health status (Green et al. 2000; Kaplan et al. 1993; Rutten van Molken et al. 1995). This has raised concerns over the ability of the VAS to elicit the strength of preference for health/disease states (Green et al. 2000). Brazier et al. (1999b) suggest that it may be that VAS techniques are capturing more of a measurement aspect of health status changes than the satisfaction or benefit conveyed by such changes as reflected by choice based methods. If this be the case, then correlating VAS and TTO values is flawed because the two are measuring different constructs, and so the low correlations cannot be used to disregard one tool in favour of the other. While the evidence from Kenya regarding correlation of the VAS and TTO is similar to previous studies, studies correlating TTO with other choice based methods (SG and PTO) and, VAS with health status measures would produce a stronger case for the argument that the two were measuring different constructs in this setting and additional evidence of validity of each.

Other constructs that were tested related to variation of values by characteristics of the respondents. Both the VAS and TTO values for the severe disease state did not vary with any of the factors. Values for the mild, moderate and very severe states however varied by gender, level of education, illness status and experience with the schistosomiasis as well as health status in the two weeks prior to the interview. Using the VAS, it appears that males were less tolerant of mild symptoms and slight effects on HRQL domains while females exhibited a more persevering attitude. Also those with no or primary education only, tended to value the mild and moderate states higher. It could be argued that these people are likely

to be poor and more susceptible to being in ill-health. Their valuation would therefore be consistent with studies that have found that those in ill health tend to provide higher valuations than healthy subjects (Badia et al. 1998; Dolan, 1996).

Variation in VAS values for the mild disease state by ratings of health status in the last two weeks contradicted previous studies (Badia et al. 1999; Boyd et al. 1990; Badia et al. 1998; Buxton and Ashby, 1988; Dolan, 1996. Read et al. 1984) while that for the moderate disease state was in agreement. This was also true for the TTO value for the very severe disease state.

Another contradiction that was noted regarded variation of TTO values for the moderate disease state by experience of illness during the valuation and experience with schistosomiasis. While those in ill health or with experience of the disease have been known to give higher values to disease states (Boyd et al. 1990; Torrance, 1987; Buxton and Ashby, 1988, Badia et al. 1999; Badia et al. 1998), those ill or with experience of schistosomiasis gave low TTO values to the moderate disease state. A possible explanation of these findings could be differences in value systems. It might be that experience of ill-health and perceived poor health status has tremendous adverse effects and implications on peoples' daily lives, such that their dislike for their states or states less than full health is intense, hence the low values. Therefore, regarding these constructs, the evidence might imply that the value systems in this community differ from those in these studies. As such, we cannot claim or disregard construct validity for the VAS and TTO based on findings from studies in other settings, but rather should seek to understand the social setting and value systems that could cause these differences.

Both the Kruskal Wallis and one way analysis of variance *F* tests suggested that there were no significant variation in values for disease states between interviewers, suggesting that there was inter-rater reliability. However, assessment of inter-rater reliability requires that the same respondent be rated twice by different interviewers on the same day. This did not happen and might bias the results of inter-rater reliability reported here and provide a basis for criticism. Therefore, future studies to improve on this should be designed so as to

minimise time between assessments and to allow for the same respondent to rate the states twice on different occasions. Such a design would also allow for inter-rater correlations, which might be a stronger statistical tool given the apparent differences between tools.

Variables relating to health and illness status and use of instruments were evaluated between test and retest to provide some evidence that would help in interpreting the test-retest coefficients for disease state values. That there were no significant differences in reports for these variables, except for the number of times the VAS procedure was explained, could be interpreted as a sign that these variables did not change. However, the correlations between test and retest valuations (see tables 8.29 and 8.30) are considered moderate as most are below 0.5 despite being significant. Rating of usual health was not expected to change, and its coefficient of 0.65 is considered low and might indicate that either people changed their perceptions of their usual health status or that they provided inconsistent answers for some reasons. These reasons though unknown may translate into inconsistent valuations, thereby biasing the reliability of the values, not because the instrument is unreliable, but because of measurement error.

In terms of values for disease states, the VAS provided reliable values for all the disease states using Spearman's correlations (r>0.57) and Kappa coefficients (0.15-0.3). However, using differences in means, only the values for mild and severe disease states were reliable. Spearman's correlation coefficients for the VAS were comparable with other studies (Kerrigan et al. 2000; Froberg and Kane, 1989b; Tan Torres, 1991). Although the Kappa coefficients are low, they are significant and represent fair agreement between test and retest (Statacorp. 2001). The evidence therefore suggests that the VAS was a valid tool for eliciting values for disease states in Kenya. The performance of the TTO in terms of reliability was slightly lower. The Spearman's correlation coefficients were below 0.5 (0.14-0.46), the suggested level for acceptable reliability (Bowling, 1997). Agreement between test and retest was only slight (Statacorp, 2001), indicating that there were changes in values between test and retest. However, differences in means indicate that values for all the disease states were reliable except the very severe state. The TTO coefficients were lower compared to other studies (Ashby et al. 1994; Sherbourne et al. 1999; Mohide et al.

1988; Dolan et al. 1996b). However, studies have also found wide variability and poor stability in TTO values (Gabriel et al. 1994; Gerard et al. 1999). In general, all tests of reliability indicated that the VAS was more reliable than the TTO, although the evidence was not very strong. Carmines and Zeller (1979) note that a low test-retest correlation may not indicate less reliability. Instead, it may signify that the underlying concept itself has changed, or due to reactivity, the fact that sometimes the very process of measuring a phenomenon can induce change in the phenomenon itself, or due to respondents misunderstanding the meaning of the task.

As test retest reliability assesses stability of values over time when there is no evidence of change (Carmines and Zeller 1979), the possibility that some factors might have changed could have caused the low reliability results for both the TTO and VAS. An examination of these factors can contribute to an understanding and interpretation of test retest reliability for these tools. Correlations for reports of illness during the valuation and two weeks prior to valuation were rather low, an indication that those reporting illness during the initial valuation were not necessarily the same during retest. Reports of illness were shown to affect values for the disease states. These differences in those reporting illness at the two time periods could have translated into genuine changes in values for disease states, thereby resulting in low coefficients. This argument holds for ratings of health status in the two weeks prior to the interviews. The correlation coefficient seems to suggest that there were changes in these reports during the two-week period between test and retest. In the case of reports of usual health status, either people revised their perceptions of their usual health status after doing the initial valuations, a process termed as reactivity (Carmines and Zeller, 1979), or their reports contained measurement error, or the respondents did not respond consistently. Other factors that could have caused values between test and retest to differ related to use of the instruments.

Table A8.19 shows factors related to use of the instruments at test and retest valuations. There was a significant improvement in the use of the VAS and a slight improvement in the use of TTO as indicated by differences in proportions between the two time periods. The correlations suggest that respondents revised their perceptions of how easy they found the

TTO and VAS. There was an increase in those finding the VAS easy and fairly easy and those finding the TTO fairly easy. In addition, in comparing the VAS and the TTO in terms of ease, there was an increase in those finding the VAS easy and those finding the TTO somewhat easy and difficult, while there was a decrease in those finding the TTO easy. Although there was an improvement in use of both tools, the TTO was relatively difficult, which might explain why the correlations for the TTO were relatively lower compared to the VAS. These changes could translate into better use of the tools and hence reduced measurement error resulting from poor understanding of the task. Hence, although the correlations would be low, the use of the instrument would actually have improved!

In general, the evidence supported reliability of both VAS and TTO values based on significance of the results. Although both techniques are considered reliable, there was stronger evidence for the VAS than TTO. However, if we consider the magnitude of the Spearman's correlation coefficients and the cut-off of 0.5 (Bowling, 1997), then only the VAS values would be considered reliable.

In order to improve on reliability of these tools, careful design of studies is called for to allow for assessment of inter-rater, intra-rater and test-retest reliability. In so doing, it would be helpful to minimise measurement error by training the interviewers adequately and ensuring that the respondents are able to use the instruments with minimum difficulties as was done in this study. It would also be useful to build in checks for consistency of responses as these could be used to infer how reliable responses are by considering variables that are expected to change or not and evaluating them. These may include socio-economic and demographic variables and variables related to use of the tools.

8.4 CONCLUSION

The VAS and TTO valuation approaches could be considered reasonably valid, reliable and practical in a rural Kenyan setting. While the majority regarded both the VAS and the TTO as practical, there was a clear preference for the VAS to TTO in which respondents had far more difficulties. VAS can therefore be considered applicable in this community while TTO would require more use and testing to gain more understanding.

Construct validity of both the VAS and TTO was argued to exist, although in some instances there were mixed findings. It was possible that the VAS and TTO measured different constructs in a disease scenario or that there were difficulties in using the methods as shown by the levels of inconsistency and wide variability in values for same disease states. More evidence on construct validity is therefore required for both instruments.

Both the VAS and TTO could be considered reliable tools for obtaining values in this setting although the evidence was modest. It was seen that there were changes in a number of variables that could have contributed to changes in values between test and retest, but also, that it was possible that there was measurement error in using the tools. However, there was stronger evidence for reliability of the VAS.

Considering the novelty of the instruments in this community, there is need for more research into the practicality, validity and reliability of the tools. As validity is paramount in a measurement tool, urgent work should test the theoretical and empirical validity of these instruments by examining the extent to which the axioms of the theories underlying the instruments are supported or violated in this setting. In addition, further construct validation should incorporate other existing valuation tools for further understanding of performance of these instruments in this setting as well as providing additional evidence to that contained in this thesis. This would guard against adopting these techniques for use in health care decision making in Kenya and elsewhere without fully ascertaining their suitability regarding their validity and reliability over time.

CHAPTER 9

DISCUSSION

9.0 INTRODUCTION

This thesis aimed to contribute to the debates surrounding the measurement and valuation of disease specific health outcomes for use in economic evaluation and health care decision making. Methodological issues in development of HRQL measures were examined through development of a new tool to assess the impact of Schistosomiasis Mansoni amongst the Kikuyu in Kenya. Issues in valuation were also examined through choice, application and examination of valuation approaches in Kenya, using Schistosomiasis Mansoni disease states. Exploration of these issues brought to the fore the cross-cultural relevance of research methods.

Drawing on the empirical findings and research literature, this chapter is devoted to discussing five main questions, each addressed in the next five sections. The first relates to how the new approach developed to assess outcomes of the impact of S. *Mansoni* on HRQL performs, in relation to other existing outcome measures.

Secondly, the issue of which valuation instrument has the strongest base for application in Kenya is examined. This question is opportune considering the growing interest in summary measures of population health e.g. QALYs (Torrance et al. 1997; Brooks et al. 2003) and DALYs (Murray et al. 2002) that require the incorporation of either disability or quality weights in their construction. Although application of these types of measures in developing countries is currently limited and mainly in research testing methodologies, there are indications that their demand in future is imminent. This is evidenced by the current widespread use of the DALY in developing countries by international agencies such as the World Bank and WHO to inform resource allocation decisions. It is also reflected in the use of, for example, the EQ-5D QALY in developing countries such as Zimbabwe.

The third question examines how well the TTO and VAS can cross cultures. With the wider use of the DALY and QALY type measures there is considerable interest in the extent of cross-cultural variation in valuations of health states (Ustun et al. 2002). It is therefore important that values elicited using different instruments are comparable across settings and this can only be ensured if equivalence in use of instruments between settings is established.

Fourth I question whether the values elicited for disease states in the thesis were representative of the impact of S. Mansoni on HRQL. Debating this question reveals how the impact of S. Mansoni on HRQL is valued, while giving a sense of the size of potential gains from investing in its control.

The fifth question is based on an illustration of using values derived for disease states to examine the potential policy implications of using a CUA for economic evaluation of a S. Mansoni intervention. This thesis did not undertake an economic evaluation of *S. mansoni*. It has however elicited disease state values that could be incorporated into CUA. This study therefore cannot make any policy recommendations regarding resource allocation, as its focus was mainly methodological issues. However, an illustrative example of a potential CUA would give indications of how different outcome measures might impact on policy. To examine the potential policy implications of using CUA for economic evaluation of a *S. mansoni* intervention, an illustrative example of cost utility analysis is presented.

9.1 Is the new way of assessing outcomes of impact of S. mansoni on HRQL better than existing outcome measures?

To debate this question, the new measure of assessing the impact of S. Mansoni on HRQL developed in this thesis is compared with three existing measures. A criteria for comparison is set out to give meaning to 'better' and to justify the usefulness and position of the new outcome measure.

The new Schistosomiasis Mansoni HRQL (SMHRQL) measurement questionnaire together with the valuation of disease states constructed from its use is compared with prevalence and intensity of infection, the existing *S. Mansoni* outcome measures. Further comparisons are made with a generic HRQL measure, using the EQ-5D (EuroQoL group, 1990, Brooks and EuroQoL group, 1996) as an example. The SMHRQL outcome measure is also compared with the DALY (Murray and Lopez, 1994, 1996) and Schistosomiasis Mansoni outcomes developed by Kirigia (1994, 1998). These instruments, with the exception of prevalence and intensity of infection, were chosen because they produce an index, are intended for use economic evaluations and have been used in a developing country setting.

The comparison criteria is based on:

- Comprehensiveness of the measure which draws on its development.
- Validity, reliability and practicality of the measure in relation to the uses the measure can be put to.
- Cross-cultural issues and generalizability of the measure.

The *S. Mansoni* HRQL (SMHRQL) instrument developed in this thesis used a clinimetric approach (Fayers and Hand, 2002). It consists of symptoms and HRQL domains and is considered to have content and construct validity and internal consistency. Disease states constructed from use of the tool were valued as a way of assessing outcomes. Kirigia (1994, 1998) developed S. Mansoni related health state descriptions representing seven main severity stages and valued them to assess outcomes. The DALY measure focuses on diseases. Its development has been heavily influenced by experts who provided severity weights for the treated and untreated sequel of different disease (Fox-Rushby, 2002). It uses disease labels and descriptions of diseases scenarios that are medically oriented. The EQ-5D is a generic measure that describes health states in terms of five HRQL dimensions (Brooks and the EuroQoL group, 1996), along with valuation sets for different countries where the instrument has been used (Brooks et al, 2003).

Table 9.1 shows a comparison of the different approaches to assessing outcomes of impact of *S. Mansoni*. Use of prevalence and intensity of infection has no information of the

impact of disease on HRQL of patients as opposed to the other outcomes. All the outcome measures considered have relied on experts to differing degrees with Kirigia (1994) and the DALY (Murray and Lopez, 1996) showing heavier reliance. Since Kirigia (1994) and the DALY (Murray and Lopez, 1996) use descriptions of disease states that are not based on any disease state descriptive system, it is difficult tell the range of states patients suffering from a given disease can fall into. The SMHRQL and the EQ-5D are however able to achieve this, in that they can be used to measure disease/health states from different populations, which are then valued to produce an index usable for economic evaluation. Because the EQ-5D relied on reviews of literature and experiences of researchers while the SMHRQL measure relied on reviews of literature, views of patients and S. Mansoni experts, the SMHRQL measure is likely to represent a more broader view of S. Mansoni and is likely to be sensitive to changes in disease states in patients, than the EQ-5D. However, this is not a fair comparison because the SMHRQL measure is disease-specific and the EQ-5D is generic and the two are likely to measure different constructs of HRQL in patients. In general, none of the instrument has started from an understanding of the local conceptions of illness and well being prior to determining the concepts to be measured as suggested in Herdman et al (1997, 1998) and Parker and Fox-Rushby (1995) and operationalized by the KENQOL Group (1996). It is obvious that Kirigia (1994) and the DALY developers (Murray and Lopez, 1994,1996) have no HRQL measurement part.

Table 9.1: Comparisons of different approaches to assessing outcomes of impact of S. Mansoni.

Outcome measures	Advantages	Disadvantages
Existing outcomes (intensity and prevalence of *S. Mansoni* infection)	• Better linked to disease and is objective	• Says nothing about impact of disease on HRQL as its based on biomedical model of disease • Cannot create an index for an economic evaluation • No information on validity and reliability of the outcomes
SMHRQL	• Incorporated views of patients and health professionals • Assesses the consequences of disease on HRQL thereby going beyond the biomedical model of disease • Disease specific hence has relevance to disease • Creates an index outcome measure for economic evaluation and has a measurement and valuation component • Instrument has validity in the Kenyan setting	• Not tested for reliability • Not tested for concurrent validity • Not in a form amenable for use in clinical settings
Values of disease states (e.g. Kirigia's and DALY)	• Creates an index outcome measure for economic evaluation.	• Not based on the views of patients, but uses 'expert' views • Not a HRQL measurement instrument. • Based on estimations of experts to compute the outcome. • No validity or reliability reported
Generic HRQL e.g. EQ-5D	• Used reviews of other measures and researchers' experience in construction. • Creates an index outcome measure for economic evaluation and has a measurement and valuation component • Instrument has validity and reliability	• Not based on fieldwork • Not validated for Kenyan setting • Mainly used in non-infectious diseases • Generic, hence might be insensitive to disease impact.

Beyond the approaches followed in developing the measures, which show whose views are incorporated, it is vital to ascertain the performance of a measure in terms of validity and reliability when put to use. The new measure was argued to have content and construct validity and internal consistency. Tests of hypotheses and correlation analysis between symptoms and HRQL indicators adduced evidence in support of construct validity for the schistosomiasis HRQL questionnaire. Although construct validity for the questionnaire was claimed based on positive correlation coefficients, the correlations were somewhat moderate, which some could interpret as weak evidence for construct validity. However, one should bear in mind that the items in this instrument were causal implying that even low correlations (with the right sign) could be interpreted as lending sufficient support for validity (Fayers and Hand, 2002).

There was no conclusive evidence regarding the relation of infection intensity to symptoms and HRQL indicators. While infection intensity is a useful indicator of morbidity, the relation of morbidity to disease may also be to past experiences rather than only current levels of worm burden (WHO, 1993). This lag may explain the apparent lack of unambiguously positive correlation between infection intensity, symptoms and HRQL domains, which implies that lack of strong evidence for construct validity does not necessarily imply that the tool is invalid. It would be interesting in future to validate infection intensity against symptoms and other HRQL measures (including generic measures) to assess the stability and or persistence of this phenomenon. This would lend information supporting or discounting use of infection intensity as an outcome measure and possibly lead to studies aimed at understanding this apparent unexpected finding.

However, lack of support for construct validity between symptoms, HRQL indicators and infection intensity might also suggest that use of infection intensity and prevalence levels, as a measure for schistosomiasis disease is a partial measure. It does not take into account the effects of suffering symptoms and their consequences on HRQL (Bowling, 2001, McDowell and Newell, 1996). On the other hand, this finding could have resulted due to existence of co-morbidities such that our measured symptoms and effects on HRQL were not entirely due to *S. mansoni*. In fact the regression results showed that presence of other

parasites in a patient was likely to result in worse off reports on all the HRQL indicators and this was especially significant for feeling of energy and strength.

Another possible explanation of the inconclusive evidence regarding the relation of infection intensity to symptoms and HRQL indicators is that this thesis took a problem centred approach that assumed that symptom experience would necessarily result in poor HRQL, rather than evaluating the patients' subjective constructions and perceptions of symptom experience on quality of life. Stenner et al (2003) note that the same measure of HRQL can take on radically different meanings to two people due to differences in their subjective constructions of quality of life. That individuals are likely to differentially interpret scale items of any HRQL measure in light of culturally available subjective constructions (Stenner et al. 2003) may explain why the subjective and objective (infection intensity) variables do not appear to relate as expected. Use of the biopsychosocial model (Engel, 1977; Truchon, 2001; Yamada et al. 2000) would have allowed for an investigation of subjective views of the patients following disease and symptom experience to gain an understanding of variation in interpretation and perception of symptoms and their effect on HRQL.

The biopsychosocial model, in contrast to the biomedical model looks at the mind and body of the patient as systems that are interlinked and treats the biological, psychological and social issues as systems of the body. The model draws a distinction between the actual pathological processes that cause disease and the patients' perception of their health and effects on it, called illness (Engel, 1977; Wikipedia, 2004). The model provides a framework for the social, psychological and behavioural dimensions of illness (Engel, 1977). Hence, two patients may be diseased, but not similarly ill, which can cause and explain variation in how subjective and objective variables measuring the HRQL constructs might fail to agree as shown in Stenner et al (2003) and explored in Yamada et al (2000) and Truchon (2001). Shweder and Bourne (1984) criticise this model of persons having a dual system of mind / body split. The assumption of a mind / body split is itself a cultural construction from the developed world which may not be transferable to a developing

country setting (Shweder and Bourne, 1984). However, this does not detract from the main idea that there may be explanations for differences based on another model.

Carmines and Zeller (1979) caution that construct validation is an on-going process and therefore it would be pre-mature to lay definite claims of construct validity of this tool with only a one-off use. Further testing with different populations, in more controlled settings and with large samples should constitute future research for further testing of validity. However, these preliminary findings show that the new SMHRQL measure has potential for being a valid tool for assessing impact of S. Mansoni disease. It performs better compared to other outcome measures because its content is justified, it has construct validity, its items internally consistent and can be used to create and index for use in economic evaluation and the outcome measure takes the HRQL concerns into consideration.

A review of studies assessing disease specific utilities in chapter four found that most studies used disease specific instruments alongside a generic instrument. This practice would be the most suitable way to assess concurrent validity for the newly developed tool. However, owing to lack of a suitable measurement instrument that could be used alongside the tool developed in this study, this form of validity could not be tested. Future studies should focus on using a combination of suitable generic instruments to facilitate better assessment of concurrent validity of the schistosomiasis tool, for example that under development by the KENQOL Group or the WHOQOL.

Two aspects that would have enhanced the validity of SMHRQL tool that were not investigated deserve mention. They relate to use of the instrument amongst children and testing of test-retest reliability of the instrument. Much of the burden from schistosomiasis falls on school-age children (Stephenson, 1993; WHO, 1993; Guyatt and Evans, 1992; Chan et al. 1994 and Evans and Guyatt, 1997) who harbor the highest prevalence and intensity of infection. Yet, due to practical difficulties associated with application of health measurement and valuation instruments to children (Bullinger, 1997; Pal, 1996; Feeny, 1999; Wright, 1996; Rosenbaum and Saigal; 1996), this study did not include those below

15 years of age (See appendix 9.1). It could therefore be claimed that the findings exclude a substantial proportion of the population (0-19 year olds comprise 58% of total population (GOK, 1997)) bearing the burden from this disease. Juniper et al. (1997), Feeny (1999), Landgraf and Abetz (1996) have shown that children over the age of 7 years are capable of responding on their own behalf and they produce reliable data. Other instruments considered for comparison with the new tool have also not been used amongst children. Future studies should therefore aim to test the extent to which children can use this tool together with valuation instruments so as to ascertain the validity of this instrument with this group. Secondly, test- retest reliability of the tool was not assessed due to time and financial considerations and should be another piece of work.

9.2 Which health state valuation technique currently has the strongest base for future use in Kenya?

To date there has been little agreement as to which method of eliciting health state values is most appropriate (Torrance, 1976; Wolfson et al. 1982, Brazier et al. 1999b, Salomon and Murray, 2002). Arguments for and against different methods have been based on ethical grounds, economic theory, and comparisons of psychometric properties (Mulkay et al. 1987; Torrance, 1976, 1986, 1987; Wolfson et al. 1982; Froberg and Kane 1989b; Brazier et al. 1999b). To debate which instrument has the strongest base for future use 'strong base' is considered in terms of practicality, validity (specifically theoretical, content, construct and concurrent) and reliability.

Suitability of a valuation technique depends on intended use of the values (Essink-Bot and Bonsel, 2002) and there are many uses (Essink-Bot and Bonsel, 2002; Fox-Rushby, 2002; Ebrahim, 1995, Patrick et al. 1993; Dolan 1997; Wilkin et al. 1992; Revicki et al. 1993; Kaplan et al. 1993). In this discussion, 'future use' relates to use in developing countries to value health benefits in economic evaluation of health care programs for priority setting and in clinical decision analysis.

The findings in this thesis supported the consensus view that the VAS performs better than the other techniques in terms of practicality (Torrance, 1976, 1986, Brazier et al. 1999b; Dolan and Sutton, 1997). The TTO was also considered to have demonstrated good practicality given that a sizeable majority understood its use quickly, found it easy to fairly easy, were comfortable using it and was the preferred technique among the patient group. Initial assessment of the VAS, TTO and SG in chapter 7 indicated that the VAS was the most practical tool followed by the TTO and SG, although the frequent mention of death in the TTO eroded its practicality drastically and was also a prominent problem affecting its appropriateness. Although the order of preference for valuation instruments was VAS, TTO and SG, a sizeable proportion preferred the TTO and SG (40% and 25%). However, they were trickier and complicated, a finding in agreement with Essink-Bot and Bonsel (2002).

The PTO is the most complex among the choice based techniques. For example, the PTO as set out in the DALY protocol (Murray, 1996) takes two days per group, each taking about 10 hours. This is a very long time for completion and requires respondents to be highly motivated to stay involved (Ustun et al. 2002; Sadana, 2002). The PTO has not been successful amongst non-health professionals (Sadana, 2002), is expensive and difficult to operationalise in a general population (Patrick et al. 1973 *cited in* Brazier et al. 1999b). These limitations clearly put the PTO at a disadvantage especially in its application amongst the general population in low-income countries where education levels are relatively lower.

In general, considering practicality shows that this study is in agreement with others that the VAS stands out followed by the TTO and SG. Despite a few problems, the VAS, TTO and SG are generally acceptability and can be easily operationalized in a general population, with relatively low levels of education when interviewer assisted.

Brazier et al. (1999) suggested economic theory as a basis for testing validity in terms of whether the assumptions of the theories underlying the different techniques correctly reflect individual's preferences. Chapter 2 showed that the SG has the strongest theoretical base despite the violation of the theoretical axioms (Brazier et al. 1999b). Other than the TTO,

which has sought theoretical support in consumer theory, the other techniques have no theoretical basis in economics. However, Brazier et al. (1999b) suggest that PTO may have its theoretical basis developed in economic literature.

While the design of this study was not geared towards explicit testing of the assumptions underlying the instruments, some inferences can be made for the TTO. That the mean values for the disease states reflected their logical sequencing in terms of severity is evidence for transitivity and consistency in choice, on average. However, individual valuations showed that there were logical inconsistencies that are violations of transitivity, although supporting the regret theory, which relaxes the transitivity axiom. As initial empirical evidence seems to lend support of transitivity using average values and violation at the individual level, theoretical validity of the TTO in the Kenyan setting remains unknown. Therefore with regard to SG, TTO, PTO and VAS, research into risk attitudes, time preference, effects of duration, distributional preferences and scaling distortions (Essink-Bot and Bonsel, 2002; Salomon and Murray, 2002) would be required in low-income settings.

There was unwillingness to trade using the TTO and SG valuation techniques, a finding common in other studies (Dolan et al. 1996; Brazier et al. 1999b), which raises serious questions about the presence of lexicographic preferences. A profound view was that if the gambling was for real, the respondent would not engage in it at all. Sadana (2002) reported unwillingness to trade off lives and reluctance to conduct any variant of the PTO in Cambodia. While the unwillingness to trade poses similar challenges for the PTO, TTO and SG, the differentiation between hypothetical and real situations confounds the practical use of the SG in Kenya.

Because all the instruments have theoretical deficiencies, consideration of the most appropriate instrument has often been based on whether the method is choice-based or choice-less. Choice-based techniques involve a trade-off and therefore incorporate sacrifice and opportunity cost. On this basis the SG, TTO and PTO have a stronger base (Brazier et al. 1999b) for future use in developing countries, all else being equal.

There is a serious deficiency in studies evaluating the content validity of valuation instrument using qualitative methods as employed in this study as shown in chapter 4. Studies reporting use of instruments in different settings, at best, compare values (Ustun et al. 2002; Salomon and Murray, 2002), rather than the concepts embodied in the instrument. While this may be indicative of the difficulties and efforts required in establishing content validity as well as the paucity of methods reported in the literature, this study deviated from that practice and using in-depth interviews assessed content validity of the VAS, TTO and SG. Claims of content validity consist of a judgement by experts regarding the comprehensiveness of an instrument in encompassing relevant content and domains to appear appropriate for the intended purpose (Streiner and Norman, 1995; McDowell and Newell, 1996). The majority of concepts and terms embodied in the VAS, TTO and SG were found to exist in Mwea. Based on this, it is argued that these three instruments have content validity, although to differing degrees given the differences in issues raised for each instrument.

The problems associated with mentioning death too often and considering perfect health unrealistic, unimaginable and impractical and difficulties expressed on placing a numerical value on a health state were common to the VAS, TTO and SG instruments. Assessment of concepts embodied in instruments is vital in understanding how they may affect values and thereby the content validity of the tool. The problems noted might introduce anchoring effects bias, i.e. whether states are valued relative to being well or to being dead, which affects values (Nord, 1992). Nord (1992) for example, reports how respondents in a study using EQ-5D VAS tended to interpret the numbers as 'percentages of fitness' implying that they were using only the upper anchor of the scale. In a more recent study, Brooks et al. (2003) pointed out that translations of EQ-5D VAS anchor points in Zimbabwe had to be changed to 'very good' and 'very bad' because people are hesitant to say their health state approximates the 'best imaginable' or the 'worst imaginable'. The way people feel about death and how they conceptualise perfect health could be relative in different places resulting in different anchoring hence the need to find out what terms and key concepts embodied in instruments mean to different people in different places. This would be helpful

in understanding, interpretation and comparison of values. While the VAS, TTO and SG are considered to have content validity, it seems that all instruments have limitations, some of which could be extended to the PTO though not tested.

Brazier et al. (1999b) noted that the literature reporting on the empirical validity of valuation techniques is not large. The VAS and TTO were considered construct valid in the Kenyan setting as mean values assigned to disease states were a decreasing function of their severity, in agreement with (Dolan, 1996). Evidence from differences in means values and correlation analysis provided weak evidence for construct validity. However, because these tools are know to give different mean values (Nord, 1992; Dolan and Sutton, 1997; Read et al. 1984; Froberg and Kane 1989b) and correlate poorly with each other (Rutten van Molken et al. 1995; Brazier et al. 1999b), these findings are not unusual and replicate those of other studies. There was also evidence that values for different disease states varied by gender, education level, illness and health status, and experience with disease in ways similar and different from previous studies. Evidence concerning the construct validity of the PTO and ME is limited (Brazier et al. 1999b) and the SG and TTO have been found to correlate reasonably well (Torrance, 1986). This means that the VAS, TTO and the SG have a stronger basis for future use in Kenya and other similar settings. However construct validation is an on-going process (Carmines and Zeller, 1979) and these instruments require further testing before definite claims of validity or lack of it can be made.

Both the VAS and TTO are generally considered to have good inter-rater reliability (Brazier et. al. 1999b) a finding supported in this study. There was stronger evidence for test retest reliability for the VAS than the TTO in terms of correlation coefficients, Kappa coefficient of agreement and tests of differences in means. Stability of values between test and retest, within and between raters is expected when there is no change between test and retest. However, it was shown that some variables such as ease of use of the instruments, illness and health status might have changed. The phenomenon of reactivity was also likely to have been at present both for the respondents and the raters. These factors might have caused the coefficients to be lower than observed in other studies especially for the TTO. Brazier et al. (1999b) building on a review by Froberg and Kane (1989b) highlight the lack

of evidence surrounding reliability of the PTO and ME. The VAS, TTO and SG all show acceptable levels of test retest reliability with higher correlations the shorter the periods as shown in table 3.4 in chapter 3. Therefore reliability of the tools favors the VAS, TTO and SG for future use. However, careful design of the studies to: minimize time between administration for inter and intra rater assessment; intensive training of interviewers for proper administration of the tools; careful demonstration and checking that respondents understand use of the tool; and better design to build in variables for consistency checks, would be helpful in enhancing reliability in future studies.

Comparing the five techniques, the VAS is best in terms of practicality. Though studies assessing the link between the VAS, TTO and SG have not found robust relationships (Brazier et al. 1999b), a breakthrough in this area could make the VAS highly suitable for future use because of its simplicity and cost saving feature. The TTO and the SG appear to have the strongest base for future use because in addition to being practical, they have demonstrated validity and reliability. They would also be suited to assessing preferences for chronic non-communicable conditions (TTO) as well as communicable conditions that carry a higher risk of immediate death (SG), because their calibrators and question framing have face validity in such areas. The PTO is weakest in terms of practicality and although little is known about its validity and reliability, it has potential for clinical decision-making and program evaluation.

9.3 To what extent are the VAS and TTO equivalent across cultures?

This section examines the extent to which VAS and TTO are equivalent across cultures. Arguments abound that the assumption of instruments being 'culture free' and that instruments developed in one culture can be applied in another is misguided (Fox-Rushby and Parker, 1995; Fox-Rushby and Parker, 1994; Herdman et al. 1997 and 1998). Such use of instruments can lead to differences in values if they end up valuing different aspects than the intended. Quality of life instruments are culture dependent or 'culture full' as quality of life is inherently shaped by and embedded in culture (Fox-Rushby and Parker, 1995).

Equivalence of instruments across cultures is considered in terms of understanding of the task (i.e. instrument-specific characteristics such as key terms and concepts embodied in the instruments that might differ across settings) and in terms of values for disease states (i.e. conceptualization of the disease states). Values for disease states could also differ between settings by characteristics of the respondents. These are discussed in turn.

Most of the concepts embodied in the VAS and TTO were found to exist in Kenya, suggesting some extent of equivalence in the valuation task. However, differences in conceptualization of death and the way it was represented in the instruments, conceptualization of perfect health and placing numeric values on health states emerged as requiring local understanding and representation, to make the instruments more culturally appropriate and ensure equivalence. These findings suggest that valuation tasks may be perceived differently, thereby limiting equivalence, if it is not ensured that the concepts that are different are described in terms that ensure equivalence. For example what constitutes being 'normal' and healthy is culture-dependent (Parker and Hopwood, 2000) as was found in this thesis. Perfect health was considered in terms of attainability and therefore requires to be described in a way that ensures equivalence to the 'perfect health' concept embodied in the instruments. Other studies (Brooks et al. 2003; Sadana, 2002) have shown that concepts like 'best imaginable', 'worst imaginable' health states require being described in local terms to ensure equivalence in valuation techniques. A finding in this thesis similar to Sadana (2002) amongst those having difficulties conceptualizing perfect health, was that telling someone they will remain in perfect health for a whole year was considered a lie. This implies that people found it difficult to believe that one could remain in a state for a whole year. This means that specification of duration that the states being valued last needs to be thought out carefully so as to appear realistic to the respondents.

The concept of death is found in both the TTO and VAS and was a major problem in the instruments and could limit equivalence in similar ways as conceptualization of 'perfect health' discussed above. Sadana (2002) in Cambodia, found that in valuing death, respondents asked "what type of death" and therefore described death as being caused by maternal death to facilitate consistent responses. In this study respondents mentioned that it

would be important to specify that one was referring to natural death, as opposed to death from super-natural causes and violence. This finding may imply two things. One is that the valuation of death could differ depending on its cause and two that the valuation of a state may differ depending on what type of death the person in that state experiences after the given duration. Circumventing the problem by replacing 'and then you die' with 'what happens after is not known' would have the effect of changing the anchor, which is also likely to affect comparability of values. This is because it is not known how respondents use 'what happens after is not known' in giving their preferences. Nevertheless the important thing is to recognize this cross-cultural difference and adjust the tools accordingly, and also to be explicit, so that those comparing values are aware of where differences are likely to arise.

The content of disease states and the importance respondents in different settings attach to such content could affect cross-cultural use of instruments. Murray et al. (2002) notes that differences in interpretation of the content of disease states and differences in expectations for domains of health and experience of symptoms that they contain could cause variation in values due cultural differences. Construction of disease states in this thesis followed after establishing content and construct validity of the measurement tool. Disease states to be valued were also chosen in terms of worsening severity implying that they represented different possible states that a schistosomiasis patient is likely to be in and were similar for all respondents. However, there was variability in values, hence it is possible that different respondents interpreted similar disease states differently and or attached differing importance to the symptoms they contained. This shows that variation does occur within an otherwise assumed homogenous group in terms of culture.

That disease states are regarded differently in different cultures was shown by Ustun et al. (2002). They reported that HIV was ranked as more disabling in Tunisia and Egypt compared to Japan, Spain and UK and attributed this to the image of HIV as a stigmatizing illness in these countries. They also found that out of 17 conditions 13 were ranked significantly differently in 14 countries. Essink-Bot and Bonsel (2002) citing Stouthard et al. (2000) state that there is evidence that it matters for valuation whether a diagnostic

disease label is used in scenarios being valued. Disease labels add information to the state to be valued, although how this information affects values has not been systematically researched (Essink-Bot and Bonsel, 2002). Brooks et al. (2003) also present information showing how EQ-5D descriptors are regarded differently in different countries (e.g. 'inability to wash and dress oneself' was regarded worse in Zimbabwe compared to UK as was 'confined to bed' and 'severe depression' in UK compared to Zimbabwe). In this thesis, there was an indication that bloody diarrhea was not considered similarly in terms of occurrence and bother by patients and experts, hence importance attached to the symptoms by both groups may differ. Valuations for scenarios containing these descriptors are likely to vary across cultures even though a similar instrument is used. This evidence shows that scenarios can be regarded quite differently and therefore comparing values without investigating the equivalence of instruments appears misguided and likely to result in costly clinical and resource allocation decisions.

Differences in interpretation of response categories across cultures can cause differences in values for states. Murray et al. (2002) demonstrates and gives examples of how individuals from different populations use categorical response scales differently. Fox-Rushby and Selai (2003) also examined how people understood the meaning of level 2 in EQ-5D and found that the middle level was not always half way between 0-100 implying differences in the conceptualization of this level. It is possible that this phenomenon could have been present in the use of the instrument in this study. During the instrument development, there were indications that meanings attached to response categories differed even across individuals in the same setting. Intensity of symptoms categories and categories of how often HRQL domains were affected may have been understood and conceptualized differently by different respondents and it is likely that these differences would persist in a different population, thereby limiting equivalence across cultures. Further research in this area would provide evidence to support cross-population comparisons of outcomes.

Human characteristics of people from different cultures are likely to affect cross-cultural use of any instrument. Important characteristics include age, gender, educational level, value system and attitude. As Brazier et al. (1999a) note there is accumulating evidence that

health state valuations vary by age, education and disease experience. In this thesis it was shown that values varied by gender, education level, illness and health status and experience with disease, in similar and different ways to other studies as discussed in chapter eight. This implies that values are likely to vary differently amongst different groups of people in same and different settings. There was expression of unwillingness to trade in using the TTO. This shows that people are likely to have different values and attitudes regarding both quantity and quality of life and their behavior in relation to time preference. Salomon and Murray (2002) conducted a study eliciting values from 69 public health professionals spanning 28 countries using the VAS, TTO, SG and PTO, where among other things they found that respondents had negative time preference in relation to TTO. Although this is contrary to conventional economic wisdom, it shows how different human characteristics can limit comparability of values.

Equivalence of valuation techniques across cultures can therefore only be claimed if equivalence is explicitly investigated and accounted for in comparing values. Describing concepts embodied in instruments in locally appropriate terms that retain equivalence can ensure equivalence of the understanding of valuation task. However, it would be difficult to ensure equivalence where differences arise from differences in values attached to disease state descriptors or human characteristics in different settings. At best, these issues need to be investigated and reported explicitly, otherwise comparisons of values across settings loses meaning.

9.4 Are the values generated for the disease states representative of the impact of *S. mansoni* on HRQL?

Four disease states were chosen to represent mild, moderate, severe and very severe disease states. While these states represented the range of states found in the sampled respondents, it probably fails to cover the entire range of possible states due to *S. mansoni*, i.e. both acute and chronic morbidity. This is so because in the sampled patients, none was found to have symptoms associated with chronic morbidity such as splenomegaly, hepatosplenomegaly and heamatemesis (WHO, 1993; Medley and Bundy, 1996; Gryseels,

1992). Therefore, the disease states measured and valued in this study are largely reflective of acute morbidity with schistosomiasis. That chronic states were not found could be due to the fact that recruitment occured at Health Center and dispensary levels whilst chronic cases are likely to be found in higher level facilities such as the District and Provincial hospitals. It could also be attributed to the sample size not being sufficiently large as well as the screening methods. Hence, further research should center on larger studies integrated at all levels of care to capture both acute and chronic morbidity.

It has been argued in this thesis that the VAS and TTO were practical, reliable and had content and construct validity and that the disease states valued had content and construct validity. This therefore suggest that the disease states that were valued were representative of acute *S. mansoni* morbidity, although this assertion is made with some caution because we found evidence of co-morbidity. An examination of differences in values for alleviating disease from the immediately worse off state to a better one (e.g. from F to E) and the symptoms and HRQL domain level associated with that are shown in table 9.2.

Table 9.2: Differences between disease states and their implied values

	From mild state to perfect health (PH-A)	From moderate to mild state (A-C)	From severe to moderate state (C-E)	From very severe to severe state (E-F)
Symptoms	• No tiredness • Can eat normally	• Somewhat tired • Can eat ½ to ¼ of normal amount of food	• Somewhat and then very tired • Can eat ½ to ¼ of normal amount of food • No watery and bloody diarrhea • No itching skin rash	• Very tired • Can eat ½ to ¼ and then no more than 2 spoonfuls of the normal amount of food • Watery and bloody diarrhea sometimes • Moderate itching skin rash
HRQL domains	Not affected any of the time	Affected a little of the time	Affected some of the time	Affected most of the time
Differences in mean Values				
VAS	0.96-0.67= 0.29	0.67-0.48=0.19	0.48-0.32=0.16	0.32-0.18=0.14
TTO	1.00-0.73= 0.27	0.73-0.55=0.18	0.55-0.24=0.31	0.24-0.00=0.24

These differences show that respondents considered alleviation of symptoms and effects on HRQL domains from moderate to mild and from mild to perfect health similarly using the VAS and TTO. However, alleviation from severe to moderate and from very severe to severe was valued differently by VAS and TTO. While the VAS gave similar values, the TTO valued these differences almost twice as much as the VAS. Considering the alleviation of symptoms and improvement in HRQL domains, the differences in values would be expected to increase as more symptoms are alleviated and HRQL domain levels moved higher. Assuming this, the differences in values using the TTO appear more realistic, since they increase as more symptoms and HRQL levels are alleviated. This would seem to enhance the face validity of the TTO values and suggest some response spreading occurred with the VAS. From this viewpoint, the TTO values appear more representative of *S. mansoni* although all other indicators suggest that the VAS values are more valid and reliable. However, as argued before it was possible that the VAS and TTO were valuing different aspects of the health state.

To give some insight into how the values of the selected *S. mansoni* disease states compare with other disease states, a comparison of VAS and TTO values obtained from other studies (Nord, 1992) is presented[68] in tables 9.3 and 9.4. Considering the disease states for which values correspond to *S. Mansoni* disease states, *S. Mansoni* disease appears to be quite severe. Use of disease labels like 'breast cancer' or 'severe angina' rather than in terms of symptoms and HRQL effects, makes the comparison at first glance portray values for *S. mansoni* disease states as exaggerating severity of *S. mansoni*. This raises the issue of whether and how use of specific disease names versus consequences of the disease in terms of symptoms and effects on HRQL would contribute to variation in values when valuing disease states (Essink-Bot and Bonsel, 2002; Ustun et al. 2002). As this issue was not investigated in this thesis, it would warrant further study.

[68] This comparison is only tentative because strictly, comparison of values can only be made after ascertaining cross-cultural equivalence of the measurement and valuation instruments used. However, this desired ideal has not been widely achieved with respect to measurement instruments (Bowden and Fox-Rushby, 2003) and is just beginning to be investigated with respect to valuation instruments (Baltussen et al. 2002, Mugo and Fox-Rushby, 2003, Fox-Rushby et al. 2000).

The age-specific disability weights for untreated and treated schistosomiasis infection are similar at 0.005 and 0.006 (QALY weight equivalent is 0.995 and 0.994) for those below and above 14 years (Murray and Lopez, 1996) which imply very mild disability from *S. Mansoni* disease. In comparison with values obtained in this study, the DALY weights for schistosomiasis appear to underrate the impact of *S. Mansoni* by representing it as a very mild condition, which is way above the mild disease state in this thesis. The value for the *S. Mansoni* mild disease state from this thesis compares with below the knee amputation and deafness weight in Murray and Lopez (1996). Kirigia's (1998) values for disease states categorized as mild to very severe are higher than those obtained in this study for each corresponding disease state. While these comparisons indicate the relative severity of *S. Mansoni,* questions can be raised about the comparability of these values as addressed in section 9.3.

Considering that the values elicited in this thesis were valid, reliable and reasonable representation of the impact of *S. mansoni* on HRQL, the above comparisons show that the impact of *S. Mansoni* is greater than implied in the schistosomiasis literature and the DALY estimations. However, the weakness and shortcomings of this study in making this claim are acknowledged.

Table 9.3: How *S. mansoni* disease states VAS values compare with other disease states

Schistosomiasis states [value]	Study*	Disease state	VAS value
Mild [0.67]	• Bombardier et al. 1982	• Needs walking stick	• 0.65
	• Llewellyn-Thomas et al. 1984	• Unable to work, some pain	• 0.68
	• Kaplan et al. 1979	• Limited walking, pain in arms and /or legs	• 0.67
	• Nord, 1991, 1992	• Unable to work, moderate pain	• 0.65
	• Kirigia, 1998	• Have bilharzia germs, mobility, self care and social participation are normal. Frequent moderate bladder and stomach pain, slight reduction in energy causing moderate reduction in capacity for livelihood activities but no absence from livelihood activities-work schooling etc.	• 0.68[a], 0.63, 0.65
Moderate [0.48]	• Bombardier et al. 1982	• Needs walking frame	• 0.47
	• Llewellyn-Thomas et al. 1984	• In house; unable to work; vomiting	• 0.48
	• Kaplan et al. 1979	• Needs wheelchair; needs help for self care; large burn	• 0.49
	• Richardson 1991	• Removed breast; stiff arm; tired; anxious; difficulties with sex	• 0.49
	• Kirigia, 1998	• Have bilharzia germs, no difficult with self care Slightly impaired mobility-can only walk for 1 mile with difficulty, persistent moderate bladder and stomach pains, moderate reduction in energy causing frequent absence from livelihood activities-work schooling; frequent absence from social community activities – church, peer get together, meetings, public *baraza* etc.	• 0.52[a], 0.46, 0.51
Severe [0.32]	• Llewellyn-Thomas et al. 1984	• In bed in hospital; needs help for self care; trouble remembering	• 0.30
	• Read et al. 1984	• Severe angina	• 0.35
	• Nord 1991, 1992	• Unable to work; limited leisure activity; moderate pain; depressed	• 0.30
	• Kirigia, 1998	• Have bilharzia germs, severely impaired mobility, bed-ridden most of the time, moderate lack of control of urination and defecation; sever reduction in energy causing total absence from livelihood activities- work, schooling; total absence from social community activities – church, peer get together, meetings, public baraza; severe body pain.	• 0.32[a], 0.28, 0.34
Very severe [0.18]	• Bombardier et al. 1982	• Needs one assistant for walking	• 0.18

* Source: Nord, 1992 in Sloan, 1996. Pp 42-3

[a]mean values from Kirigia (1998) are from health professionals, teachers and farmers respectively.

Table 9.4: How *S. mansoni* disease states TTO values compare with other disease states

Schistosomiasis states [value]	Study	Disease state	TTO value
Mild [0.73]	• Buxton et al. 1987	• Breast cancer: removed part of breast; occasionally concerned	• 0.72
		• Removed breast: occasionally concerned	• 0.70
Moderate [0.55]	• Bombardier et al. 1982	• Needs walking frame	• 0.58
	• Read et al. 1984	• Severe angina	• 0.53
Severe [0.24]	• Bombardier et al. 1982	• Needs one assistant for walking	• 0.28
	• Buxton et al. 1987	• Removed part of breast; stiffness of arm; engulfed by fear; unable to meet people	• 0.27
Very severe [0.00]		• Dead[a]	• 0.0

* Source: Nord, 1992 in Sloan, 1996. Pp. 42-3 [a] This is not normally valued in TTO but is assumed to be 0.

9.5 What are the potential policy implications of using CUA for economic evaluation of a *S. mansoni* intervention?

The illustration uses as a template, a previous cost effectiveness study (Carabin et al. 2000), but computes QALYs based on values obtained from this thesis. Cost per QALY is then compared with cost per infected patient treated obtained in Carabin et al. (2000). Potential policy implications are discussed based on these findings. This illustration starts with some background information, followed by estimation of QALYs, comparison of results and discussion.

A Medline search for studies on cost effectiveness and cost utility analysis of *S. mansoni* intervention strategies in Kenya found seven studies in Africa, of which one was in Kenya (Kirigia, 1997). All except Carabin et al (2000) focused on *S. heamatobium* rather than *S. mansoni*. Carabin et al. (2000) was also the first study to assess the cost effectiveness of alternative screening strategies in delivering treatment for *S. mansoni* in a Primary Health Care Centre (PHCC) setting, and was therefore the most appropriate study to base this illustration on. Carabin et al. (2000) has the features of a mass population chemotherapy and selective chemotherapy, which would be consistent with the approach taken in this

thesis, which did not target any group and whose patient recruitment was at the health care setting.

Carabin et al. (2000) considered three chemotherapy strategies for schistosomiasis morbidity control in Burundi based on health care delivery through 17 primary health care centres. The first strategy involved treating all patients with symptoms suggestive of *S. mansoni*. All symptomatic patients were treated using praziquantel. The second strategy was screening with Kato-katz, where all patients were screened and those testing positive were treated with praziquantel at 40mg/kg of body weight. The third strategy was passive case detection using blood in stool as an indicator for infection. All patients reporting blood in stool were treated with praziquantel. In total 41,051[69] patients visited the PHCCs. Through screening, 3892[70] infected patients were treated while the passive case detection using blood in stool only had 316[71] infected patients treated. Further details of resource requirements for each strategy, epidemiological parameters and cost computations are provided in Carabin et al. (2000).

Carabin et al (2000) used the number of infected persons treated, as the outcome measure and found the Kato-Katz smear to confirm diagnosis as the most cost effective option. To adjust the outcome measure to QALYs for this illustration, estimates of probabilities[72] of being in different *S. Mansoni* disease states are required and these were obtained from Kirigia (1994, 1997), and are shown in table 9.5. They were however adjusted to exclude normal and comatose states (table 9.6). To adjust the probabilities, it was assumed that those who did not visit the health facility were in the normal state. Further, we assume that there were none in comatose, to be consistent with values obtained in this thesis. This means that the outcomes computed will be underestimated because they do not take the chronic morbidity into account. Hence the percentages in Kirigia's estimates are adjusted

[69] Using the first strategy, 92.52% of patients were treated but did not have the infection.
[70] This strategy treated only those who were infected, therefore did not miss any.
[71] This strategy missed to treat 92% of infected patients, which implies that case detection using blood in stools left out infected patients.
[72] Kirigia (1994) used three experts (2 local, 1 foreign) to estimate the probabilities with which various disease states would be distributed. The values presented here refer to such probabilities assuming the status quo option. This involves treatment when one visits the health facility.

so that the percentage accounted for by those treated using each strategy in Carabin et al. (2000) across disease states is 100.

Table 9.5: Estimates of distribution of health states assuming status quo

State	Health state prevalence (%)*	Health state prevalence (%)**
Normal	44	25
Mild	22	30
Moderate	22	25
Severe	8	11
Very severe	4	6
comatose	0	3
Totals	100	100

Source: *Kirigia, 1994. ** These were estimated from Fig. 2, Kirigia, 1997 as they appeared slightly different from those in Kirigia 1994.

Table 9.6: Adjusted Health State Prevalence rates (%)*

Disease states	1994	1997
Mild	39.3	41.7
Moderate	39.3	34.7
Severe	14.3	15.3
Very severe	7.1	8.3
	100	100

* Adjusted to account for only those who present at health care facilities

All costing information was based on Carabin et al. (2000) since they considered *S. mansoni* control strategies. The study was thus a source of information on total economic costs and estimated number of patients treated for each strategy (table 9.8). The number of patients treated for each strategy was distributed across the disease states as shown in table 9.9, using the probabilities in table 9.6. Using the VAS and TTO disease state values (table 9.7), QALYs for each option were computed (table 9.10). Table 9.11 compares the cost per QALY gained using both the VAS and TTO values with cost per infected patient treated (from Carabin et al. 2000) for the three control strategies.

Table 9.7: Disease state values

Disease states	VAS	TTO
Mild	0.67	0.73
Moderate	0.48	0.55
Severe	0.32	0.24
Very severe	0.18	0

Table 9.8: Estimates of costs and number of infected patients treated by strategy

	Control Strategy		
	Kato	Symptoms	Blood in Stool
Economic costs (US$)	16,351.66	48,375.10	1,141.65
Estimated number of infected patients treated	3,892	41,051	316

Source: Carabin et al. 2000. Table 3. Pp.196.

Table 9.9: Distribution of patients treated across the four disease states by control strategy.

Disease state	Using Kirigia 1994 estimates			Using Kirigia 1997 estimates		
	Kato	Symptoms	Blood in stool	Kato	Symptoms	Blood in stool
Mild	1529	16127	124	1622	17105	132
Moderate	1529	16127	124	1351	14254	110
Severe	556	5864	45	595	6272	48
Very severe	278	2932	23	324	3421	26
Totals	3892	41051	316	3892	41051	316

This illustration has implicitly made a number of assumptions, some of which may not be realistic, but which suffice for the demonstration. It assumes that after treatment, patients in different states are returned to perfect health. The intervention is evaluated for one year only, therefore there is no discounting of either the costs or the benefits. For simplicity and consistency with cost estimates, all epidemiological information and economic parameters for Burundi are assumed to be similar to those for Kenya. The VAS and TTO values are taken to be valid and reliable representations of community preferences. It is also assumed

that in the absence of health state distribution data from randomised controlled trials, the probabilities from Kirigia (1997 and 1994) are correct.

Table 9.10 shows that treating all patients with symptoms yielded the most QALYs, followed by screening and treating and lastly, treating only those reporting blood in stool. There are slightly more QALYs gained using VAS values than using TTO values. This ordering of interventions in terms of effectiveness outcome was in agreement with Carabin et al. (2000). The outcome measure 'number of patients treated' used in Carabin et al (2000) was as expected higher than the QALYs gained arising from the differences in the nature of outcome measure used. The use of number of patient treated assumes equal benefits from treatment irrespective of the severity of disease state a patient is alleviated from, i.e assigns a weight of one to all states infected patients are in. The use of QALY recognises that different patients are in different disease states and takes into account the consequences of being in each state on health related quality of life, through the quality weight assigned to each disease state.

Table 9.10: QALYs gained by intervention strategy using VAS and TTO values

	Estimate of total QALYs gained by each option and valuation method			
	VAS-1994*	VAS-1997**	TTO-1994	TTO-1997
Kato	1906	1908	1801	1822
Symptoms	20100	20126	19001	19220
Blood in Stool	155	155	146	148

* Refers to Kirigia, 1994 estimates. ** Refers to Kirigia, 1997 estimates

Economic costs for each strategy were used to compute the cost per QALY gained. As reported in Carabin et al. (2000) treating all patients with symptoms was the most expensive, followed by screening and treating, while the option of treating only those with blood in stool was the least costly. However, this option misses out 92% of infected patients. Treating all patients with symptoms was very costly as it treated 90% non-infected patients. Therefore these two options have some shortcomings as observed by Carabin et al (2000).

Table 9.11 presents the potential cost utility and cost effectiveness ratios. Clearly, the results are very different. The most costly option, treating all patients with symptoms suggestive of *S. mansoni*, in Carabin et al.(2002) is the cheapest using CUA, irrespective of the valuation technique, although VAS CUA ratios are slightly lower. This option is nearly five times cost effective compared to Carabin et al (2000). Treating those reporting blood in stool is second cost effective while screening is the least cost effective. The cost effectiveness ratios in the latter two are about twice as high as those of Carabin et al. (2000) are. Hence, the ranking of options obtained in Carabin et al. is quite different from that obtained using CUA in terms of cost per QALY gained.

Table 9.11: Cost per QALY gained by valuation method and intervention (US$)

Illustration using QALYs	**Kato**	**Symptoms**	**Blood in stool**
VAS-1994	8.58	2.41	7.38
VAS-1997	8.57	2.40	7.37
TTO-1994	9.08	2.55	7.81
TTO-1997	8.97	2.52	7.72
Carabin et al. 2000. Using number of infected patients			
Cost per infected patient treated	4.20	12.43	3.61

Considering the cost per QALY gained and the percentage of patients, who would be left untreated using reports of blood in stool, this option appears less desirable compared to screening which though slightly costly, treats all infected patients. Where the prevalence rates are high, the option of treating all with symptoms suggestive of *S. mansoni* would appear as the cheapest strategy because fewer non-infected people would be treated. One could speculate that this strategy is akin to mass population chemotherapy, suggesting that in cases of high prevalence it would be a cost-effective strategy for morbidity control, considering that fewer uninfected people would be treated. Carabin et al. (2000) concluded that screening for S. *mansoni* was the most cost effective, findings consistent with Guyatt et al. (1994) regarding *S. haematobium* in Tanzania, but at variance with those of this illustration.

The differences in conclusions noted using different outcome measures imply that resource allocation prioritisation decisions and policies regarding choice of control strategies based on the two would be fundamentally different. Assuming a given budget level, in the case of Carabin et al. (2000), screening for infection would be the recommended policy, whereas using the CUA illustration, mass treatment based on symptoms would be the recommended policy, despite the fact that all other parameters remain the same except the outcome measure. The conclusion reached using the CUA illustration here, reflect those of Kirigia (1997) where mass and selective treatments using praziquantel had the greatest health improvements.

As Dunlop (1984) notes, there is considerable miss-specification of benefits in economics of parasitic disease reduction, in that they are not even conceptually considered, let alone empirically measured and valued. This can lead to fewer resources allocated to control of such diseases. This is a relevant policy issue considering that previous studies have not incorporated HRQL gains in benefit identification, measurement and valuation, as it potentially implies that *S. mansoni* control efforts may have been under-funded in the past. This thesis has shown that an important part of specification of benefits from schistosomiasis interventions has previously not been accounted for and that this mis-specification of benefits could potentially bias resource allocation and policy decisions. This lack of understanding of the potential gains from reduction and control of this disease may partly explain the apparent lack of enthusiasm, attention and focus on this disease from researchers and donors and low prioritisation by ministry of health policy makers. Further research towards refinement of methods of outcome measurement from *S. mansoni* intervention strategies would provide invaluable information to policy makers that would aid better resource allocation decision making. A move towards outcome measures such as the QALY and the DALY would better reflect the consequences of infections such as *S. Mansoni* on people's HRQL and also contribute towards appropriate resource allocation. As Crompton et al. (2003) observe, "*the result of a re-calculation of the DALY due to schistosomiasis is eagerly awaited. Much depends on the quality of data and the disability weighting assigned to the disease. What we now call schistosomiasis has been recognised*

as a dreadful affliction throughout our history" (pp. 123). This thesis has made its contribution as one amongst the few venturing into this task of making the burden due to *S. mansoni* better understood.

However, this thesis measured and valued the impact of *S. mansoni* using non-monetary techniques. A comparison with the monetary valuation through the willingness to pay approach could also shed light on the subject and render the findings useful for CBA analysis. Future research into willingness to pay would be useful as it could also point towards how sustainable intervention programs could be given that cost sharing in GOK health care delivery system in Kenya has been implemented with varying degree of success and consequences for access to health care. As Guyatt (2003) notes a national de-worming program in Kenya would require substantial investments. With a GNP per capita of US$ 360 (World Bank, 2003), this kind of investment may be unaffordable for Kenya and alternative sources of funding may need to be explored. Hence, willingness to pay studies would inform this policy issue that CUA studies have limitations addressing.

CHAPTER 10

CONCLUSIONS

The broad aim of this thesis was to contribute to the debates surrounding the measurement and valuation of disease specific health outcomes for use in economic evaluation and health care decision making. *S. Mansoni* disease was used a case study to illustrate the debates. The conclusions are presented in three parts. The first is devoted to a summary of main findings from each chapter. The second presents the methodological and empirical contributions to knowledge. Finally, by drawing on lessons learnt, suggestions for future research are made.

10.1 Summary of thesis

Preferences (values or utilities) for health/disease states form an important component in the construction of health outcomes for use in economic evaluation and health care decision making. Welfare theory, consumer theory under certainty and theories of consumer choice under risk and uncertainty were identified as the relevant theoretical basis for preference elicitation in chapter two. These theories were considered to have a role in understanding consumer choices, valuations of health and health care and facilitation of isolation of risk attitudes, and hence were considered complimentary in aiding descriptions of consumer behavior in making choices.

Chapter three explored issues of both measurement and valuation of health, paying special attention to developing countries. While noting that most studies do not make the distinction between HRQL measurement and valuation, the need for a distinction was underscored to facilitate exploration of economic efficiency. It was shown that to date both HRQL measurement and valuation instruments have been used largely in Europe and North America where they were largely developed for chronic conditions. A few such instruments are used in developing countries or amongst parasitic diseases such as *S. Mansoni*, and

hence their performance in these settings is unknown and considered inappropriate. Five non-monetary valuation techniques were assessed in terms of their validity, reliability and practicality. The VAS, TTO and SG have been used and researched more extensively than the PTO and ME, and their performance suggested they were more likely to be suitable for testing in Kenyan setting.

Chapter four critically and systematically reviewed the methodological and empirical issues relating to the measurement and valuation of disease states. The review revealed that when used in new settings, transfer of instruments was not always guided by considerations of validity and reliability, but ease of use and popularity. The majority of studies reviewed assessed construct validity, although it has been noted that assessment of content validity should precede any other form of validity assessment. There was virtually no testing of content or criterion validity of instruments when used to assess disease specific utilities (DSU). A few studies assessed reliability and practicality of instruments when used to assess DSU. When assessed, indicators of practicality and analytical tools for reliability were not applied uniformly across instruments causing difficulties in comparisons of performance of instruments.

Initial reviews of the schistosomiasis literature revealed lack of a suitable HRQL instruments that could be used to measure disease/ health states arising from illness with *S. Mansoni*. In addition, no valuation work had been undertaken with respect to *S. Mansoni* disease states as previous studies undertaking economic evaluation of *S. Mansoni* had focused on intermediate outcome measures that failed to account for quality of life lived in different morbidity states. In chapter five, through consultation of literature and involvement of patients and health professionals, a content valid HRQL instrument was developed to assess the impact of *S. Mansoni* on HRQL using the clinimetric approach (Fayers and Hand, 2002). Consultation of literature identified 16 symptoms and 8 HRQL domains constituting the long form questionnaire, which through consultation with patient and health professionals was reduced to 11 symptoms and 6 HRQL domains. The 11 symptoms were abdominal pain and discomfort, diarrhoea, watery diarrhoea, bloody diarrhoea, tiredness, nausea, loss of appetite, itching skin rash, dizziness, fever and

vomiting. The 6 HRQL domains included mobility, performance of daily duties, performance and output of work, feeling of energy and strength, social participation and feeling of worry and anxiety. While there was more agreement between patients and experts on symptoms and HRQL domains, a striking finding was the discordance regarding the importance of bloody diarrhea, watery diarrhea and diarrhea, raising concerns about the prior use of diarrhea as an indicator of *S. Mansoni*.

Use of the questionnaire amongst patients and community members provided evidence supporting construct validity of the tool, as reported in chapter six. The questionnaire differentiated between patients and community members in terms of age, marital status, occupation, reports of illness, knowledge and experience of *S. Mansoni*. Patients tended to be younger and single and more reported illness, health problems, and had experienced *S. Mansoni*. Significantly more patients experienced symptoms for a longer duration. Construct validity was also ascertained between symptoms and HRQL domains. Although the correlations were below 0.5, all those above 0.22 were statistically significant. As Fayers and Hand (2002) suggest, when an instrument contains causal variables, as did this one, even low correlations can indicate validity. Evidence of correlations between infection intensity and symptoms was mixed and somewhat weak ($r<0.18$), suggesting a complex relation between infection intensity and symptoms and HRQL domains. There was no evidence for construct validity between infection intensity and HRQL domains. Eight symptoms and five HRQL domains adduced evidence of construct validity on one or more occasion, and were therefore considered to be valid items for incorporation into a measure. The symptoms included abdominal pain and discomfort, tiredness, nausea, fever, loss of appetite, dizziness, itching skin rash and bloody diarrhoea. The HRQL domains included mobility, performance of daily duties, feeling of energy and strength, social participation and feeling of worry/anxiety. The questionnaire was considered generalizable to other settings, subject to assessment of equivalence of response categories and expression of different symptoms as some were found to require locally appropriate expressions.

Construction of disease state scenarios and testing content validity of VAS, TTO and SG valuation techniques in Mwea was covered in chapter seven. The disease states were

considered valid representation of *S. Mansoni* as they were based on the questionnaire whose content and construct validity were precisely ascertained. Content validity for the VAS, TTO and SG entailed ascertaining existence of concepts and key terms embodied in each instrument as well as their appropriateness and practicality in a Kenyan rural community.

Most concepts and key terms embodied in the instruments existed in Mwea in a similar way to the intention of the techniques. However, there were problems with some key concepts such as 'perfect health' and representation of 'death'. For example, 18% of the respondents regarded perfect health as unrealistic, unimaginable and impractical to think of in their setting and so it was rated well below 1.0. The concept was also considered in terms of attainability, which has implications for cross-cultural comparison of values and results of economic evaluation, as values would need re-scaling before valid and meaningful comparisons could be made. 38% of the respondents felt that death was mentioned too often. This contributed to the inappropriateness of the TTO and SG for this setting. Modifications were therefore required to account for locally appropriate approaches to 'death talk'.

Unwillingness to trade was also a problem with both the VAS and TTO, which raises serious questions about the functioning of the TTO and SG, which depend on the existence of lexicographic preferences. It also raises questions of whether these views are likely to differ by culture and or perceived risk of mortality. Such problems highlighted the inappropriateness of comparisons of values across cultures, without assessing their conceptual equivalence first. Assessing conceptual equivalence would help; in making necessary modifications of the instruments; to improve the relevance of instruments to local cultures, and; to understand the extent to which comparisons of values across countries is valid and meaningful. This thesis, like previous studies, found that the VAS was regarded better than the TTO and SG in terms of practicality, both of which were more or less similar. A general contention in this thesis is that the VAS, TTO and the SG are all content valid and can be used with rural Kenyan populations, although none of the instruments were problem free. However, the problems that were noted were important pointers to the

importance of assessing conceptual equivalence of the instruments in new settings, and further evidence of how 'culture-full' instruments are. The VAS performed best on all aspects considered, with the TTO and SG performing similarly.

Chapter eight reported on the practicality, validity and reliability of the VAS and TTO for use in a rural Kenyan population using *S. mansoni* disease states. The VAS and TTO valuation approaches were considered reasonably valid, reliable and practical in this setting. While both the VAS and the TTO were considered sufficiently practical, there was a clear preference for the VAS by respondents who had far more difficulties with the TTO. The VAS was therefore considered applicable in this community with TTO requiring further testing prior to further use. The evidence regarding the construct validity of the VAS and TTO, was somewhat mixed. Ranking of disease states using mean values for groups as a whole followed a logical sequence with the worse off disease states assigned lower values using both techniques. This finding suggests that the valuation given to a disease state is a decreasing function of its severity (Dolan, 1996) and supports construct validity for both the TTO and the VAS. There was a higher number of logical inconsistency using the TTO, although it has been argued (Dolan and Kind, 1996; Loomes and Sugden, 1982) that this need not be considered as representing non-transitive preferences. There was more variability in the TTO values compared to VAS, although the VAS showed some evidence of response spreading.

Differences in mean TTO and VAS values were statistically significant, with TTO values for the mild and moderate states higher than VAS values and vice versa for the severe and very severe states. The correlations for the severe and very severe states were positive, low and statistically insignificant. The VAS and TTO were considered reliable tools for obtaining values in this setting, although the evidence was modest. It was seen that there were changes in a number of variables that could have contributed to changes in values between test and retest, but also that it was possible that there was measurement error in using the tools.

Chapter nine draws the findings of the empirical study together in a discussion addressing five questions regarding the aims of the thesis. The discussion showed that the new approach to assessing the outcomes of *S. Mansoni* on HRQL was considered better than the existing approaches based on its content and construct validity and that it could be used to create an index for use in economic evaluation. However, further construct validity and reliability is required in future studies. There was stronger evidence of the practicality, construct validity and reliability of the VAS compared to the TTO. However, the TTO and SG were considered content valid and both are choice-based and have theoretical basis in economics. It was therefore argued that the VAS, TTO and SG have potential for future use in Kenya and other similar settings. However, it was shown that health measurement and valuation instruments require to be assessed for equivalence in different cultural settings before meaningful comparisons of outcomes can be made. The discussion also argued that the values for disease states elicited in this study were valid and reliable representations of impact of *S. Mansoni* on HRQL and that this impact is greater than currently implied in the schistosomiasis literature and DALY estimations. Based on a CUA illustration, it was shown that use of intermediate outcome measures results in mis-specification of benefits from treatment and can potentially result in inappropriate resource allocation decisions.

10.2 Contribution to knowledge

This thesis has made a number of methodological and empirical contributions to knowledge. Methodological contributions are discussed first as the thesis is deemed to have made a more significant contribution in this area.

Methodological contribution

Reviews of literature have contributed to knowledge through clarification of issues related to measurement and valuation of health outcomes in the process of QALY construction. The need to separate measurement and valuation of health outcomes and establish clear links between the two steps was highlighted and followed in this thesis, thereby contributing towards promoting good practice in eliciting disease specific utilities. In

particular, the reviews helped in identification of knowledge gaps, some of which contributed to the design and scope of this thesis as well as identifying the methods.

The review of studies eliciting disease specific utilities was a contribution to knowledge. Prior to this thesis such a systematic review had not been attempted. It highlighted the limited testing of validity, reliability and practicality of instruments. There appears to be an implicit assumption that the instruments are equivalent everywhere, an assumption challenged by findings of this thesis and previous literature Herdman et al. (1998). The thesis revealed the paucity of use of both HRQL measurement and valuation instruments in developing countries and therefore engaged in measurement and valuation of S. Mansoni disease states in its empirical part.

The development of a *S. Mansoni* HRQL tool was novel and important contribution of this thesis. Steps were taken to ensure that the views of both patients and schistosomiasis experts were taken into account in the development of the tool thereby ensuring it content validity. Previously, measurement of outcomes of *S. Mansoni* infections had been restricted to prevalence and intensity. Further, measures of outcomes for economic evaluation have been intermediate measures such as cases treated, reduction in intensity, reduction in prevalence and life years saved. By developing the HRQL tool, this thesis has ventured into a previously unexplored possibility of measuring the impact of *S. Mansoni* on HRQL. This work has provided a basis for researchers wishing to replicate or further develop this tool and to economists and policy makers wishing to incorporate the concerns of quality of life lived with *S. Mansoni* infection into their outcome measures to aid decision making. The tool provides a *S. Mansoni* disease state classification system and would therefore be useful in: measuring the extent of the disease and its impact on HRQL in a given population; measuring changes in disease states that can be valued to construct outcome measures for economic evaluation; and also in monitoring patients to facilitate clinical decision making.

This thesis contributes to those wishing to develop disease specific instruments in developing countries. By incorporating the views of patients and health professionals after choosing symptoms and HRQL domains from literature, a valid disease specific instrument

that assesses the impact of disease on HRQL could be developed. However, this approach had weakness that future disease specific instrument developers should consider. To ensure a solid foundation, upon which performance of an instrument can be judged, it would be helpful to begin with qualitative research with focus groups of patients, before consulting health professionals and the literature. Eliciting patients' views regarding how the disease manifests in terms of symptoms, how and which HRQL areas it effects and the importance of such symptoms and domains to them would ensure content relevance of the instrument and provide a basis for developing constructs to validate. It would ensure that terms used in the instrument are locally generated, thereby improving its practicality.

The use of the VAS, TTO and SG together in an African setting, was novel. The qualitative approach and careful steps taken in testing content validity of these techniques, when they were applied in settings other than North America and Europe, provided new knowledge as the content validity of valuation techniques had not been evaluated in this manner hitherto. An understanding of how key terms and concepts embodied in these instruments were conceptualized in the Kenyan setting was an addition to knowledge. This understanding highlighted what modifications were deemed appropriate to maximize conceptual equivalence and also challenged notions that valuation instruments are universal or 'culture free'. It was shown that differences in conceptualization of some concepts and key term embodied in the valuation instruments could result in erroneous comparisons of values if these cultural considerations are not accounted for. The work in this thesis has helped to advance knowledge in the new and growing literature (Fox-Rushby and Parker, 1995; Herdman et al, 1997, 1998; Brooks et al. 2003; Murray et al. 2002) highlighting the importance of accounting for culture, in the measurement and valuation of health outcomes. Only when cross-cultural equivalence of instruments is ensured can meaningful comparisons of health outcomes across populations and cultures be made.

In testing for construct validity of valuation techniques, the use of different approaches that included ranking using mean values, testing for differences between means and correlation analysis produced information indicating that different methods do not always arrive at the same conclusion. Future studies assessing construct validity of valuation techniques would

find this information useful and incorporate it in their choice of methods. Finally, the finding that valuation approaches developed in North America and Europe can be used in a rural Kenyan setting with subjects of relatively low educational level (provided modifications were made to account for cultural differences), was new knowledge. This was particularly important for the TTO and SG, as neither had been previously used in Kenya.

Empirical contribution

The work contained in this thesis is the first attempt to measure and value the impact of *S. Mansoni* on HRQL amongst patient and community members. This provided important information about the range of symptoms that the respondents associated with *S. Mansoni* infections, how these symptoms disrupted their daily duties and their impact on HRQL. Use of the measurement tool showed that *S. Mansoni* has adverse effects in terms of symptom experience and their effects on HRQL of those infected. This was new information as prior to this work, no study had embarked on assessing the impact of S. Mansoni disease on HRQL, and reports of outcomes from suffering the disease were restricted to prevalence and intensity of infection. Empirical findings also questioned the use of bloody diarrhoea, as an indicator of *S. Mansoni* as the symptom was not reported frequently within a sample of known patients. It was shown that a number of symptoms could be associated with *S. Mansoni* infection, although there was a possibility of confounding from co-morbidities.

Further empirical contribution was from the findings on validity, reliability and practicality of the valuation techniques. The VAS, TTO and SG were shown to have practicality and content validity in the Kenyan setting, although requiring modification to account for culture. In addition, initial evidence of the construct validity of the VAS and TTO was demonstrated in valuing *S. Mansoni* disease states. They were also both regarded as having practicality and sufficient reliability with their first application in this setting to warrant future use. The range of values indicated that S. Mansoni disease states have severe adverse effects on people's HRQL, an issue that has received little attention in previous studies (Guyatt, 1998. 2003; Guyatt and Evans, 1992, 1995; Guyatt et al. 1995, 1998) that have

tended to use intermediate outcomes that do not incorporate paient preferences. Although more work is still required in this area, this study has demonstrated that the impact of S. Mansoni on HRQL is far more than currently acknowledged. This finding is relevant for other diseases, especially in developing countries, where the impact of the majority of diseases on HRQL remains unknown.

Contribution to potential policy implications

Taking the assessment of outcomes of *S. Mansoni* beyond prevalence and infection intensity opens up the possibilities of using the technique of cost utility analysis in decision making concerning *S. Mansoni* interventions within the health care sector, based on evidence rather than the estimations of 'experts'. Through the CUA illustration, this thesis exemplified the potential for inappropriate resource allocation decisions when existing outcome measures are used. However, until actual CUA studies are undertaken this potential policy implication remains empirical and can only be taken as an indication of future prospects. Nevertheless, the findings in this thesis suggest that future economic evaluation studies on *S. Mansoni* should incorporate HRQL concerns as exemplified herein, so that the basis of resource allocation decisions and policy making is guided more explicitly and with evidence.

10.3 Suggestions for future research

This thesis has highlighted the scarcity of studies on measurement and valuation of health in developing countries. It is crucial that methodological work developing new instruments and determining the equivalence of adapted instruments and testing their performance in new settings, both in measurement and valuation of health be undertaken in developing regions. This will prepare a firm foundation for use of such instruments and facilitate cross population comparisons of health and disease as well as economic evaluation results. This is in view of the imminent demand for such outcome measures as resource allocation decisions veer towards the use of cost/DALY averted.

There is need for improvement of the SMHRQL tool developed in this study in future. Such improvements should entail revising the SMHRQL questionnaire to include only the symptom and HRQL domains that were valid. The improvements should be informed by additional studies seeking focus groups input from patients, before seeking consulting literature and health professional to strengthen the basis for content validity and to inform future adaptation to other settings. Patients should be involved in deciding which symptoms and HRQL domains are important to them as well as deciding how important they are as this may determine/affect future valuations of disease states. It would also help to include a wide cross-section of patients covering all possible disease states, as this study found that those with chronic S. Mansoni states were not identified at the health center. Therefore, in future recruitment of patients should be at both lower and higher level facilitates to ensure fair coverage. Additionally, co-morbidities need to be studied more carefully as there are possibilities of confounding which may affect performance of the instruments. In studying co-morbidities researchers should be aware of the difficulties of working in resource poor environments that have few facilities to screen or treat patients for the whole range of potential co-morbidities. The sample of patients and health professionals in this thesis was small. Future studies should aim at larger samples. The final instrument should be shorter to be usable in clinical settings.

Together with improving on the current disease measurement tool, more work is required in testing validity and reliability of the instrument in Kenya and other settings in developing countries. Findings regarding the importance of bloody diarrhea as a symptom indicative of *S. Mansoni* disease were challenged in this thesis and it would be important to find out if this was unique to Mwea or if it would happen in other settings where S. Mansoni is prevalent in Kenya and beyond. Results of construct validity failed to adduce support that infection intensity, HRQL domains and symptoms were correlated as expected. Therefore infection intensity did not seem to relate to immediate symptom experience or effects on HRQL domains. Further studies are needed to confirm or challenge this rather surprising result and attempt to unravel the mechanisms at work.

More work on construct validation of the measurement tool and with a sample of patients representing all the disease states would be useful, as it was found that only acute morbidity was represented in the sample on this thesis. Of interest in this pursuit would be the exploration construct validation using combination of generic instruments such as the KENQOL that has been undergoing development in Kenya (Bowden, 2001). Also, following proper adaptation procedures, the possibility of transferring measurement instruments such as the EQ-5D, SF-6D, HUI would widen the possibilities for construct validation of the tool as well as validation of the KENQOL, when it is ready for use. This points to possibilities of future work in cross-cultural adaptation of HRQL instruments in developing country settings. In doing this, it would help to follow the guidelines of Herdman et al. (1998). Reliability of the SMHRQL tool was not assessed and hence future studies should be designed to undertake testing of different forms of reliability.

Although the VAS and TTO were found to be valid, reliable and practical in the Kenyan setting, their application was quite novel. More studies in developing country settings would therefore confirm or dispute these findings in line with the contention that validation is an on-going process. Such studies should aim to test further the usefulness of the SG in this setting because it has the strongest theoretical base in economics, is choice based and it incorporates risk and uncertainty in its formulation, which resemble decision making in health care.

To date no studies have examined the theoretical validity of valuation tools in developing country settings, and this should be part of future studies. In particular, studies of risk attitudes, time preference, measurement scale properties of the instrument calibrators and distributional concerns in various settings would help in understanding why and how values are likely to differ in these settings. At the same time, systematic studies of factors causing variation in values would be required as this forms vital information for decision-makers. Also, to further inform on the issue of 'whose values' and how they differ between groups, comparative valuation should be undertaken amongst health professionals, other professionals in the general population and policy makers, in addition to patients and general population.

The assessment of reliability of the VAS, TTO and SG valuation instruments in this thesis was not as thorough as possible. Future studies should assess inter-rater, intra-rater reliability that require administration of the tool preferably the same day and they have not been widely reported on. In assessing test retest reliability, studies should ensure that the time lapse between administration is long enough for subjects not to remember their responses and short enough for variables to have not changed. This is a delicate balance, however should change be anticipated, it would be helpful to build in checks in terms of variables to assess if there are changes that could explain differences between test and retest.

The highest burden from *S. Mansoni* is on school-age children (5-19 year olds), yet the cut off for this thesis was 15 years. This means that the views of a substantial proportion of those who bear the burden from this illness are still not known. With due appreciation of the difficulties involved in using health measurement and valuation instruments with children, future studies should aim to test the extent to which children can use these instruments and what modifications would need to be made to make them more suitable and usable with children.

Despite all the shortcomings in this thesis for which future research to improve on is suggested, the thesis has generated important findings. It has shown that S. Mansoni has adverse impacts on HRQL through symptom experience and their effects on HRQL domains. It has also shown that patients and members of the general population consider these impacts to be quite severe based on the valuations they assigned to some selected disease states. The thesis has also shown that the VAS, TTO and SG could be used to elicit valid and reliable values for disease states in a developing country rural setting where respondents have relatively low levels of education. Underscored in this work is the importance of assessing and accounting for cross-cultural equivalence of both measurement and valuation tools in new settings to enable meaningful comparisons of population health and values used in economic evaluations.

REFERENCES

Abdel Wahab, MF and Strickland GT. (1993) Abdominal ultrasonography for assessing morbidity from schistosomiasis. 2. Hospital studies. *Trans R Soc Trop Med Hyg 87(2): 135-7.*

Ackerman, SJ, Rein, AL, Blute, M, Beusterien, K, Sullivan, EN, Tanio, CP, Manyak, MJ and Strauss, MJ (2000) Cost Effectiveness of Microwave Thermotherapy in Patients with Benign Prostatic hyperlasia: Part 1-Methods. *Urology 56: 972-980*

Adams, EJ, Stephenson, LS, Latham, MC and Kinoti, SN (1994) Physical Activity and Growth of Kenyan School Children with Hookworm, *Trichuris Trichuria* and *Ascaris lumbricoides* infections Are improved after Treatment with Albendazole. *Journal of Nutrition , 124:1199-1206.*

Allen, T, Amuyunzu, M., et al. (1997) "Conceptions of Health in Health Related Quality of Life." *Quality of Life Research 6(7/8): 614.*

Amuyunzu, M, Allen T, et al. (1995) "The Resonance of Language: Health Terms in Kenya." *Quality of Life Research 5(4): 428-429.*

Anderson, RM and Medley, GF (1985) Community Control of Helminth Infections of Man by Mass and Selective Chemotherapy. *Parasitology 90: 629-660.*

Apajasalo, M, Rautonen, J. et al. (1996) "Quality of Life in Pre-adolescence: A 17-Dimensional Health Related Measure (17-D)." *Quality of Life Research 5: 532-538.*

Arpinelli, F, De Carli, G, Irvine, SH, Bamfi, F and Recchia, G (1995) Constructing a Health-Related Quality of Life Questionnaire for Italian Peptic Patients. *Quality of Life Research 4: 391.*

Ashby, J, O'Hanlon, M and Buxton MJ. (1994) "The Time Trade Off Technique: How do the Valuations of Breast Cancer Patients Compare to Those of Other groups?" *Quality of Life Research 3: 257-265.*

Avis, NE and Smith, KW (1995) The Development of the MQOL-HIV: A Quality of Life Instrument for persons with HIV. *Quality of Life Research 4: 392.*

Awasthi, S, Bundy, DAP and Savioli, L (2003) Helminthic Infections. *British Medical Journal, 327:431-3.*

Badia, X, Herdman, M et al. (1998) The Influence of Ill-Health Experience on the valuation of Health. *Pharmacoeconomics 13(6): 687-696.*

Badia, X, Roset, M et al. (1999) "Inconsistent responses in Three Preference-Elicitation Methods for Health States." *Social Science and medicine, 49: 943-950.*

Baker, R and Robinson, A. Responses to Standard Gambles: Are Preferences "well Constructed"? Health economics, 2004, 13(1): 37-48.

Bakker, C, Rutten, M, van Santen-Hoeufft, M, Bolwijn, P, van Doorslaer, E, Bennett, K and van der Linden, S (1995) Patient Utilities in Fibromyalgia and the Association with Oter Outcome Measures. *Journal of Rheumatology 22:1536-43.*

Baltussen, RMPM, Sanon, M, Sommerfeld, J and Wurthwein, R (2002) Obtaining Disability Weights in Rural Burkina Faso using a Culturally Adapted Visual Analog Scale. *Health Economics 11(2): 155-164.*

Bayoumi, AM and Redelmeier, DA (1999) Economic Methods for Measuring the Quality of Life Associated with HIV Infection. *Quality of Life Research 8: 471-480.*

Bell, DE (1985) Disappointment in Decision-Making Under Uncertainty. *Operations Research 33(1):1-27.*

Bennett, KJ, Torrance, GW, Boyle, MH and Guscott, R (2000) Cost-Utility Analysis in Depression: The McSad Utility Measure for Depression Health States. *Psychiatric Services 51: 1171-1176.*

Berhe N, Gundersen, SG, Abebe, F, Birrie, H Medhin, G and Gemetchu, T (1999) Praziquantel side effects and efficacy related to Schistosoma mansoni egg loads and morbidity in primary school children in north-east Ethiopia. *Acta Trop. 72(1):53-63.*

Bernard, G (1986) A Discussion of the Present State of Utility Theory. *(FUR 11, VENICE, 1984) Theory and Decision (1986) 22:173-188.*

Binmore, K (1999) "Why Experiment in Economics?" *The Economic Journal 109(453): F16-F24.*

Birch, S and Gafni, A (1992) Cost Effectiveness / Utility Analyses: Do Current Decision Rules Lead Us to Where We Want to be? *Journal of Health Economics. 11: 279-296.*

Biswas, T (1997) Decision-making Under Uncertainty Macmillan Press LTD. *Chapters 1 and 5.*

Blaxter, M and Paterson, E (1982) Mothers and Daughters: A Three Generational Study of Health Attitudes and Behaviour. London: Heinemann.

Bleichrodt, H. A new Explanation for the Difference between Time Trade-Off Utilities and Standard Gamble Utilities. *Health Economics, 2002, 11: 447-456.*

Bleichrodt, H and Johannesson, M. An Experimental Test of a Theoretical Foundation for Rating Scale Valuations." *Medical decision Making,* 1997; *17: 208-216.*

Blumenschein, K and Johannesson, M (1998) Relationship Between Quality of Life Instruments, Health State Utilities, and Willingness to Pay in Patients With Asthma. *Annals of Allergy, Asthma and Immunology 80: 189-94.*

Bombardier, C, Ware, J, Russell, IJ, Larson, M, Chalmers, A and Read, JL (1986) Auranofin Therapy and Quality of Life in Patients with Rheumatoid Arthritis: Results of a Multi-Center Trial. *The American Journal of Medicine, 81(4): 565-578.*

Bonner, J (1986) Politics, Economics and Welfare: An Elementary Introduction to Social Choice. Harvester Press.

Borooah, VK (2001). Logit and Probit: Ordered and Multinomial Models. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-138. Thousand Oaks, CA: Sage.

Bosch, JL and Hunink, MGM (1996) The Relationship Between Descriptive and Valuational Quality-of-Life Measures in Patients with Intermittent Claudication. *Medical Decision Making 16: 217-225.*

Bowden A (2001) Variation in Health-Related Quality of Life: A Case Study of an Akamba Population. Ph.D Thesis. School of Geography, University of Manchester.

Bowden, A and Fox-Rushby, JA (2000) A Critical review of the Adaptation and Use of Generic HRQL measures Amongst Populations Outside of North America, Western/Southern/Northern Europe, Australia and New Zealand. *Quality of Life Research, 9(3): 313.*

Bowden, A and Fox-Rushby, JA (2003) A Systematic and Critical Review of the Process of Translation and Adaptation of generic Health-related Quality of Life Measures in Africa, Asia, Eastern Europe, the Middle East, South America. *Social Science and Medicine, 57:1289-1306.*

Bowden, A, Fox-Rushby, JA, Nyandieka, L and Wanjau, J (2002) Methods for Pre-testing and Piloting Survey Questions: Illustrations from the KENQOL Survey of Health-Related Quality of Life. *Health Policy and Planning, 17(3): 322-330.*

Bowling, A (1997) Measuring Health: A Review of Quality of Life Measurement Scales. *Open University Press.*

Bowling, A (2001) Measuring Disease: A Review of Disease-Specific Quality of Life Measurement Scales. *Second edition, Open University Press.*

Boyd, NF, Sutherland, HJ et al. (1990) "Whose Utilities for Decision Analysis?" *Medical Decision Making* 10: 58-67.

Brandstatter, E and Kriz, WC (2001) Hedonic Intensity of Disappointment and Elation. *The Journal of Psychology, 135(4):368-380.*

Brazier, J and Deverill, M (1999) A Checklist for Judging Preference-Based Measures of Health Related Quality of Life: Learning from Psychometrics. *Health Economics 8: 41-51.*

Brazier, J and Dixon, S (1995) The Use of Condition Specific Outcome Measures in Economic Appraisal. *Health Economics 4:255-264.*

Brazier, J, Deverill, M and Green, C (1999a) A Review of the Use of health Status Measures in Economic Evaluation. *Journal of Health Services Research and Policy, 4(3): 174-184.*

Brazier, J, Deverill, M, Green, C, Harper, R, and Booth, A (1999b) A Review of the Use of Health Status Measures in Economic Evaluation: *Health Technology Assessment 3(9).*

Brazier, JE, Harper, R, Munro, J, Walters, SJ and Snaith, ML (1999c) Generic and Condition-Specific Outcome Measures for People with Osteoarthritis of the Knee. *Rheumatology. 38: 870-877.*

Brazier, JE, Usherwood, TP, Harper, R, Thomas, K (1998) Deriving a Preference Based Single Index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology. 51(11):1115-29.*

Brooker, S, Miguel, EA, Waswa, P, Namunyu, R, Moulin, S, Guyatt, H and Bundy, DA (2001) The Potential of Rapid Screening Methods for Schistosoma Mansoni in Western Kenya. *Annals of Tropical Medicine and Parasitology. 95(4):343-51.*

Brooks, R (1996) EuroQoL: The Current State of Play. *Health Policy 37: 53-72.*

Brooks, R, Coons, JS, de Cook, P, Devlin, N, de Weerdt, W, Hansen, P. Ikegami, N, Jelsma, J, Johnson, J and Marra, C (2003) EQ-5D in Selected Countries around the World. *In eds. Brooks R., Rabin R. and de Charro, F. (2003) The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective.*

Brooks, R, Rabin R and de Charro, F (2003) The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective.

Brooks, RG (1995) Health Status Measurement; A Perspective on Change. Macmillan.

Broome, J (1993) QALYs. *Journal of Public Economics 50:149-167.*

Broome, J (1996) The Value of Life and the Value of Population. *Journal of Population economics 9:3-18.*

Brown, GC (1999) Vision and Quality of Life *Transactions of American Ophthalmological Society. 97: 473-511.*

Brown, GC, Brown, MM, Sharma, S, Brown, H, Gozum, M and Denton, P (2000) Quality of Life Associated with Diadetes mellitus in an Adult Population. *Journal of Diabetes complications. 14: 18-24*

Buckingham, JK, J, Birdsall, et al. (1996) "Comparing three versions of the time trade off: time for a change?" *Med Decis Making 16(4): 335-47.*

Bullinger, M, Anderson, R et al. (1993) "Developing and Evaluating Cross-Cultural Instruments from Minimum Requirements to Optimal Models." *Quality of Life Research 2: 451-459.*

Bullinger, M. (1997) The challenge of international quality of life research. *Psychology and Health, 1997, 12, 815-825.*

Butterworth, AE, Curry, AJ et al. (1994) "Immunity and morbidity in human schistosomiasis mansoni." *Trop Geogr Med 46(4 Spec No): 197-208.*

Butterworth, A. E., Sturrock, R.F., Ouma, J.H., Mbugua, G.G., Fulford, A.J., Kariuki, H.C. and Koech, D. (1991). "Comparison of Different Chemotherapy Strategies Against *Schistosoma Mansoni* in Machakos District ,Kenya: Effects of Human Infection and Morbidity." *Parasitology, 103,: 339-355.*

Buxton, M and Ashby, J (1988) The Time Trade Off Approach to Health State Valuation. Measuring Health: A Practical Approach. T. G. Smith, John Wiley and Sons Ltd.

Cairns, J (1996) Measuring Health Outcomes: Condition Specific and Patient specific Measures are of Limited Use when Allocating Resources. *BMJ 313: 6.*

Cano, SJ (2001) Development and Validation of Patient-Based Measures of Outcome in Plastic Surgery. Ph.D. Thesis. London School of Hygiene and Tropical Medicine, University of London. U.K.

Carabin, H, Guyatt, H and Engels, D (2000) A Comparative Analysis of the Cost-Effectiveness of Treatment Based on Parasitological and Symptomatic Screening for Schistosoma Mansoni in Burundi. *Tropical Medicine and International Health. 5(3): 192-202.*

Carmines, EG and RA Zeller (1979) Reliability and Validity Assessment. Sage Publications.

Carr-Hill, RA (1989) Assumption of the QALY Procedure: Background Material for the workshop on QALYs. *Social Science and Medicine 29(3):469-477.*

Carr-Hill, RA (1992) A Second Opinion: Health Related Quality of Life Measurement-Euro Style. *Health Policy, 20:321-328.*

Chan, MS, Guyatt, HL, Bundy, DAP and Medley, GF (1994) "The Development and Validation of an Age-structured Model for the Evaluation of Disease Control Strategies for Intestinal Helminths." *Parasitology. 109: 389-396.*

Chan, MS, HL, Guyatt, et al. (1995) "The development of an age structured model for schistosomiasis transmission dynamics and control and its validation for Schistosoma mansoni." *Epidemiol Infect 115(2): 325-44.*

Chan, MS, Medley GF, Jamison, D and Bundy, DAP (1994) "The Evaluation of Potential Global Morbidity Attributable to Intestinal Nematode Infections." *Parasitology. 190: 373-387.*

Chancellor, JVM, Coyle, D and Drummond, MF (1997) Constructing Health State Preference Values from Descriptive Quality of Life Outcomes: Mission Impossible? *Quality of Life Research 6: 159-168.*

Chandiwana, SK and NO Christensen (1988) "Analysis of the dynamics of transmission of human schistosomiasis in the highveld region of Zimbabwe. A review." *Trop Med Parasitol 39(3): 187-93.*

Chandiwana, SK, P Taylor, et al. (1988) "Prevalence and Intensity of Schistosomiasis in Two Rural Areas in Zimbabwe and their relationship to Village Location and Snail Infection Rates." *Annals of Tropical Medicine and Parasitology 82(2): 163-173*

Chen, AY, Daley, J and Thibault, GE (1996) Angina Patient's Ratings of Current Health and Health without Angina: Associations with Severity of Angina and Comorbidity. *Medical Decision Making. 16: 169-177.*

Chen, S, Shaheen, A and Garber, A (1998) Cost Effectiveness and Cost-Benefit Analysis of Using Methotrexate Vs Geockerman Therapy for Psoriasis: A Pilot Study. *Archives of Dermatology. 134: 1602-1808.*

Cohen, B (1996) "Is Expected Utility Theory Normative for Medical Decision Making?" *Medical Decision Making 16: 1-6.*

Cohen, B (1996) "Reply: Utilitarianism, Risk Aversion and Expected Utility." *Medical Decision Making, 16: 14*

Connolly, KJ and Kvalsvig, JD (1993) Infection, Nutrition and Cognitive Performance in Children. *Parasitology, 107:S187-S200.*

Crompton, DWT, Engels, D, Montresor, MP, and Savioli, NL (2003) Action Starts Now to Control Disease due to schistosomiasis and Soil-transmitted Helminthiasis. *Acta Tropica. 86: 121-124*

Dinwiddy, C and Teal, F (1996) Principles of Cost Benefit Analysis for Developing Countries. Cambridge University Press.

Dolan, P (1996) "Modelling Valuations for Health States: The Effect of Duration." *Health Policy 38: 189-203.*

Dolan, P (1996) "The Effect of Experience of Illness on Health State Valuations." *J Clin Epidemiol 49(5): 551-564.*

Dolan, P (1997) Modeling Valuations for EuroQoL Health States. *Medical Care. 35(11): 1095-1108.*

Dolan, P and Kind, P (1996) Inconsistency and Health State Valuations. *Soc. Sci. Med. 42(4): 609-615.*

Dolan, P, C, Gudex, et al. (1996a) "Valuing Health States: A Comparison of Methods." *Journal of Health Economics 15: 209-231.*

Dolan, P, C, Gudex, et al. (1996b) "The time trade-off method: results from a general population study." *Health Econ 5(2): 141-54.*

Dolan, P. and M. Sutton (1997) "Mapping visual analogue scale health state valuations onto standard gamble and time trade-off values." *Soc Sci Med 44(10): 1519-30.*

Donaldson, C (1998) The (Near) Equivalence of Cost Effectiveness and Cost Benefit Analysis: Fact or Fallacy? *Pharmacoeconomics 13(4):389-396.*

Douzdjian, V, Ferrara, D and Silvestri, G (1998) Treatment Strategies for Insulin-Dependent Diabetics with ESRD: Cost-Effectiveness Decision Analysis Model. *American Journal of Kidney Diseases. 31(5): 794-802.*

Dranitsaris, G, Leung, P, Mather, J and Oza, A (2000) Cost-Utility Analysis of Second Line Hormonal Therapy in Advanced Breast Cancer: A Comparison of Two Aromatase Inhibitors to Megestrol Acetate. *Anti-Cancer Drugs 11:591-601.*

Drummond, MF, O'Brien, BJ, Stoddart, GL and Torrance, GW (1997) Methods for the Economic Evaluation of Health Care Programmes.

Dunlop, DW (1984) Theoretical and Empirical Issues in benefit Identification, measurement and Valuation Related to Parasitic Disease Control in Poor Countries. *Social Science and Medicine 19(10): 1031-37.*

Ebrahim, S (1995) Clinical and Public Health Perspectives and Applications of Health Related Quality of Life Measurement. *Social Science and Medicine. Vol. 41 No. 10: 1383-1394.*

Eiser, C (1997) "Children's Quality of Life Measures." *Arch-Ids-Child 77(4): 350-4.*

Eiser, C, T Havermans, et al. (1995). "Development of a Measure to Assess the Perceived Illness Experience after Treatment for Cancer." *Archives of Diseases in Childhood 72: 302-307.*

Engel GL The Need for a new Medical Model. Science, 1977, 196: 129-136
Nunnally J and Bernstein I. Psychometric Theory. 3rd Edition. New York: McGraw Hill; 1994

Essink Bot, ML, GJ, Bonsel, et al. (1990) "Valuation of health states by the general public: feasibility of a standardised measurement procedure." *Soc Sci Med 31(11): 1201-6.*

Essink-Bot, M and Bonsel, GJ (2002) How to Derive Disability Weights. In eds. *Murray CJL, Salomon, JA, Mathers, CD and Lopez, AD 2002. Summary Measures of Population Health: Concepts, Ethics, measurement and Applications. WHO, Geneva.*

EuroQol Group (1990) "A New Facility for the Measurement of Health Related Quality of Life." *Health Policy 16: 199-208.*

Evans, DB and HL Guyatt (1997) "Human Behaviour, Cost-Effectiveness Analysis and Research and Development Priorities: The Case of Schistosomiasis vaccine." *Tropical medicine and International Health 2(11): A47-A54.*

Fayers, PM and Hand, DJ. (2002) Causal Variables, Indicator Variables and Measurement Scales: An Example from Quality of Life. *Journal of the Royal Statistical Society. Series A., 165(2):233-261.*

Feeny, D (1999) "Why Not Ask the Kids Themselves?: Using Health-related Quality-of-life Measures in Paediatric Studies." *Pharmaco-economic News.*

Fink. A (1993) Evaluation Fundamentals: Guiding Health Programs, Research, and Policy. Sage publications.

Fishburn, P and Wakker, P (1995) The Invention of the Independence Condition for Preferences. *Management Science 41(7):1130-1144.*

Fishburn, PC (1989) Non-Linear Preference and Utility Theory., Wheatsheaf Press, Brighton. Book Review.

Fox-Rushby, J (1993) Measurement and Valuation of Health Related Quality of Life: The case of Differing Pacing Modalities. PhD. Dessertation, Economics. London, Brunel University.

Fox-Rushby, J (1994) The Relationship Between Health Economics and Health Related Quality of Life." In Orley, J. Kuyken, W. (eds) *Quality of Life Assessmen: International Perspectives. Spinger-Verlay.*

Fox-Rushby, J (1995) "Questioning Premises: Health Related Quality of life in Kenya." Quality of Life Research 4: 428.
Fox-Rushby, J (1996) "Introduction." *Health Economics and Financing Exchange.* (No.9 Winter).

Fox-Rushby, J (For the KENQOL Group) (2000) Operationalising Conceptions of 'Health' amongst the Wakamba and Maragoli of Kenya: The Basis of the KENQOL Instrument. *Quality of Life Research 9(3): 316.*

Fox-Rushby, J and M Parker (1994) " Accounting for Culture in the Measurement of Health Related Quality of Life". *Quality of Life Research 3: 41-42.*

Fox-Rushby, J and Parker, M (1995) Culture and the Measurement of Health-Related Quality of Life. *European Review of Applied Psychology 45(4):257-263.*

Fox-Rushby, J and Selai, C (2003) What Concepts Does the EQ-5D Measure? Intentions and Interpretations. *In eds. Brooks R., Rabin R. and de Charro, F. (2003) The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective.*

Fox-Rushby, J, K, Johnston, et al. (1997) "Creating an Instrument to Assess Lay Perceptions of HRQL: Options and Implications." *Quality of Life Research 6(7/8): 633.*

Fox-Rushby, JA (2002). Disability Adjusted Life Years (DALYs) for Decision-Making? An Overview of the Literature. Office of Health Economics. UK.

Fox-Rushby, JA, Bowden, A, Nyandieka, L and Wanjau, J (2001) The Reliability and Validity of a new Generic Measure of Health-related Quality of Life for use amongst the Kamba of Kenya. Paper Presented at IHEA 3rd World Congress, York, UK.

Fox-Rushby, JA, Nzioka, C, Nganda, B and Mugo, M (2000) Are Health State Valuation Techniques Likely to be Understood in the same way in Kenya? Moving Towards an Assessment of Conceptual Equivalence. Background Paper presented at Workshop 'Assessing HRQL – What can the Cognitive Science Contribute? Hull University.

Fox-Rushby, JA, Nzioka, C, Nganda, B and Mugo, M (2001) How well do Health Valuation Techniques Cross Cultures? Evidence from Kenya. Paper Presented at IHEA 3rd World Congress, York, UK.

Freemantle, N (1996) Are Decisions taken by Health Care Professsionals Rational? A Non Systematic Review of Experimental and Quasi Experimental Literature. Health Policy, 38:71-81.

Frisch, D and Clemen, RT (1994) Beyond Expected Utility: Rethinking Behavioral Decision Research. *Psychological Bulletin, 116(1):46-54.*

Froberg, DG and RL Kane (1989b) " Methodology for Measuring Health State Preferences-11: Scaling Methods.1." *J. Clin Epidemiol 42(5): 459-47.*

Froberg, DG and RL Kane (1989c) "Methodology for Measuring Health State Preferences-111: Population and Context Effects." *J. Clin Epidemiol 42(6): 585-592.*

Froberg, DG and RL Kane (1989d) "Methodology for Measuring Health State Preferences-1V: Progress and a Research Agenda." *J. Clin Epidemiol 42(7): 675-685.*

Furlong, W, D, Feeny, et al. (1990) Guide to Design and Development of Health State Utility Instrumentation. Hamilton Ontario. McMaster University.

Furlong, W, D, Feeny, et al. (1998) Multiplicative Multi-Attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System: A Technical Report. Hamilton, Ontario, McMaster University.

Gabriel, S.E., Kneeland, TS, Melton, JL, Moncur, BE and Tosteson, ANA (1999) Health-Related Quality of Life in Economic Evaluations for Osteoporosis: Whose Values Should we Use? *Medical Decision Making. 1999, 19: 141-148.*

Gabriel, SE, Campion, ME and O'Fallon, WM (1994) A Cost-Utility Analysis of Misoprostol Prophylaxis for Rheumatoid Arthritis Patients Receiving Nonsteroidal Antiinflamatory Drugs. *Arthritis and Rheumatism, 37(3): 333-341.*

Gafni, A (1994) The Standard Gamble Method: What is Being Measured and How It Is Interpreted. *Health Services Research; 29(2):207-224.*

Gerard, K, Johnston, K and Brown, J (1999) The Role of a Pre-Scored Multi-Attribute Health Classification Measure in Validating Condition-Specific Health State Descriptions. *Health Economics. 1999, 8: 685-699.*

Giesler, RB, Ashton, CM, Brody, B, Bryne, MM, Cook, K, Geraci, JM, Hanita, M, Souchek, J and Wray, NP (1999) Assessing the Performance of Utility Techniques in the Absence of a Gold Standard. *Medical Care, 1999. 37(6): 580-588.*

Girman, C, Hartmaier, S, Thiboutot, D, Johnson, J, Barber, B, deMuro-Mercon, C and Waldstreicher, J (1995) Development of an Acne-Specific Quality of Life Questionnaire. *Quality of Life Research, 4: 433.*

Glahe, FR and Lee, DR (1981) Microeconomics: Theory and Applications. HBJ. *Chapters 4 and 5.*

GOK (1989) 1989 census report vol. 1 and 2.

GOK (1994) Statistical Abstract, 1994. Nairobi, CBS.

GOK (1995) Statistical Abstract, 1995. Nairobi, CBS.

GOK (1996) Statistical Abstract, 1996. Nairobi, CBS.

GOK (1997) Development Plan: 1997-2001. Nairobi, Ministry of Planning and National Development.

GOK (1997) Kirinyaga District Development Plan, 1997-2001. Nairobi, Ministry of Planning and National Development.

Gold, MR, Seigel, JE, Russel, LB and Weinstein, MC (1996) Cost Effectiveness in Health and Medicine. Oxford University Press.

Green, C, Brazier, J and Deverill, M (2000) Valuing Health-Related Quality of Life: A Review of Health State Valuation Techniques. *Pharmacoeconomics, 17(2): 151-165.*

Gryseels, B (1989) The relevance of schistosomiasis for public health. *Trop Med Parasitol 40(2): 134-42.*

Gryseels, B (1992) Morbidity due to infection with Schistosoma mansoni: an update. *Trop Geogr Med 44(3): 189-200.*

Gryseels, B and L Nkulikyinka (1989) Two-year follow-up of Schistosoma mansoni infection and morbidity after treatment with different regimens of oxamniquine and praziquantel. *Trans R Soc Trop Med Hyg 83(2): 219-28.*

Gryseels, B and Nkulikyinka, L (1989) Two-year follow-up of Schistosoma mansoni infection and morbidity after treatment with different regimens of oxamniquine and praziquantel. *Trans R Soc Trop Med Hyg 83(2): 219-28.*

Guilleman, F, Bombardier, C and Beaton, D (1993) Cross-Cultural Adaptation of Health Related Quality of Life Measures: Literature Review and Proposed Guidelines. *Journal of clinical Epidemiology. 1993, 46(12): 1417-1432.*

Guyatt, GH (1995) The Philosophy of Health-Related Quality of Life Translation. In eds. Shumaker SA and Berzon, R. The International Assessment of Health Related Quality of Life: Theory, Translation, Measurement and Analysis. *Rapid communications of Oxford Ltd.*

Guyatt, GH, DH Feeny, et al. (1993) "Measuring Health Related Quality of Life." *Annals of Internal Medicine. 118: 622-629.*

Guyatt, GH, King DR, Feeny, DH, Stubbing, D, and Goldstein, RS (1999) Generic and Specific Measurement of Health Related Quality of Life in a Clinical Trial of Respiratory Rehabilitation. *Journal of Clinical Epidemiology, 52(3): 187-192.*

Guyatt, H (2003) The Cost of Delivering and Sustaining a Control Programme for Schistosomiasis and Soil-transmitted Helminthiasis. *Acta Tropica, 86: 267-274*

Guyatt, HL (1998) Different Approaches to Modeling the Cost-Effectiveness of Schistosomiasis Control. *Mem-Inst-Oswaldo-Cruz 93(suppl.1):75-84.*

Guyatt, HL and D, Evans (1995) "Desirable characteristics of a schistosomiasis vaccine: some implications of a cost effectiveness analysis." *Acta Tropica 59: 197-209.*

Guyatt, HL and Evans, D (1992) Economic Considerations for Helminth Control. Parasitology Today 8:397-402

Guyatt, HL and MS, Chan (1998) An Investigation into the Interaction between Drug Efficacy and Drug Price of Praziquantel in Determining the Cost Effectiveness of School Targeted Treatment for *Schistosoma mansoni* Using a Population Model. *Tropical medicine and International Health. 3(6): 425-435.*

Guyatt, HL, Chan, MS, medley, GF, and Bundy, DAP (1995) Control of *Ascaris* Infection by Chemotherapy: Which is the Most Cost-effective Option? *Transactions of the Royal Society of Tropical medicine and Hygiene 89:16-20.*

Guyatt, HL, D Evans, et al. (1994) "Controlling schistosomiasis: the cost effectiveness of alternative delivery strategies." *Health Policy and Planning 9(4): 385-395.*

Guyatt, HL, Bundy, DAP et al. (1993) A Population Dynamic Approach to the Cost Effectiveness Analysis of Mass Antihelminthic Treatment: Effects of Treatment Frequency on Ascaris Infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene 87: 570-575.*

Guyatt, HL, Chan, MS, et al. (1995) "Control of Ascaris infection by chemotherapy: Hagen, O (1985) Rules of Behaviour and Expected utility Theory. Compatibility versus Dependence. *Theory and Decision, 1985; 18: 31-45.*

Hadju, V, Stephenson, LS, Abadi, K et al. (1996) Improvements in Appetite and Growth in Helminth-Infected School Boys Three and Seven Weeks after a Single Dose of Pyrantel Pamoate. *Parasitology, 113:497-504.*

Hall, J, Gerard, K, Salkeld, G and Richardson J (1992) A Cost Utility Analysis of Mammography Screening in Australia. *Social Science and Medicine, 1992. 34(9): 993-1004.*

Hatz, C, Savioli, L, et al. (1990) " Measurement of Schistosomiasis-Related Morbidity at Community Level in Areas of Different Endemicity." *Bulletin of the World Health Organisation 68(6): 777-787.*

Havranek, EP, Mcgovern, KM, Weinberger, J, Brocato, A, Lowes, BD and Abraham, WT (1999) Patient Preferences fore Heart Failure Treatment: Utilities are Valid Measures of

Health Related Quality of Life in Heart Failure. *Journal of Cardiac Failure. 1999. 5(2): 85- 91.*

Hayman, AJA, Fairclough, DL, Harris, JR, and Weeks, JC (1997) Patient Preferences Concerning the Trade-Off between the Risks and Benefits of Routine radiation Therapy after Conservative Surgery for Early Stage Breast Cancer. *Journal of Clinical Oncology, 1997. 15:1252-1260.*

Herdman, M, Fox-Rushby, J and Badia, X (1997) 'Equivalence' and the Translation and Adaptation of Health Related Quality of Life Questionnaires. *Quality of Life Research, 6:237-247.*

Herdman, M, Fox-Rushby, J, Rabin, R, Badia, X and Selai, C (2003) Producing Other Language Versions of the EQ-5D. *In eds. Brooks R., Rabin R. and de Charro, F. (2003) The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective.*

Herdman, M, Fox-Rushby J and Badia, X (1998) A Model of Equivalence in the Cultural Adaptation of HRQoL Instruments: The Universalist Approach." *Quality of Life Research 7: 323-335.*

Hey, JD. (1979) Uncertainty in Microeconomics. Martin Robertson. *Chapter 4.*

Holland CV (1987) Hookworm Infection. In eds. Stephenson, LS (1987) Impact of HelminthvInfection on Human Nutrition. London: taylor and Francis, pp. 128-160.

Hornby, AS (1995) Oxford Advanced Learner's Dictionary of Current English. Fifth edition. Oxford University Press.

Hunt, SM (1991) The Problem of Quality of Life. *Quality of Life Research, 6: 205-212.*

Hunter, JM, Rey, L, et al. (1993) Parasitic diseases in water resources development the need for intersectoral negotiation, Geneva World Health Organisation 1993.

Hurny, C, van Wegberg, B et al. (1998) Subjective health estimations (SHE) in patients with advanced breast cancer: an adapted utility concept for clinical trials. *Br J Cancer 77(6): 985-91.*

Jacobson, JJ, Maxson, BB, Mays, K and Kowalski, J (1992) A utility Analysis of Dental Implants. *International Journal of Maxillofacial Implants. 1992. 7: 381-388.*

Jaoko, WG, G, Muchemi, et al. (1996) "Praziquantel side effects during treatment of Schistosoma mansoni infected pupils in Kibwezi, Kenya." *East Afr Med J 73(8): 499-501.*

Jelsma J (2002) The Burden of Disease Due to Disability in a High Density Area of Harare, Zimbabwe. Doctoral Thesis, Katholieke Universiteit Leuven, Boimedical Sciences Faculty of Physical Education and Therapy, Leuven.

Jelsma J, Chivaura, VG, de Weerdt, W, and de Cook, P (2000) A Bridge Between Cultures: A Report on the Process of Translating the EQ-5D Instrument into Shona. *South African Journal of Physiotherapy, 2000; 56(4): 3-9.*

Jelsma J, Chivaura, VG, Mhundwa, K, de Weerdt, W, and de Cook P (2000) The Global Burden of Disease Disability Weights. *The Lancet, 355 (9220): 2079-2080.*

Jelsma J, Shumba, D, Hansen, K, de Weerdt, W and de Cook P (2002) Preferences of Urban Zimbabweans for Health and Life Lived at Different Ages. *Bulletin of the World Health Organisation, 2002. 80(3): 204-209.*

Johannesson, M (1996) Theory and Methods of Economic Evaluation of Health Care. Kluwer Academic Publishers.

Johnston, K, J, Brown, et al. (1998) Valuing temporary and chronic health states associated with breast screening. *Soc Sci Med 47(2): 213-22.*

Jones A (2001) Applied Econometrics for Health Economists – A Practical Guide. Office of Health Economics.UK

Juniper, EF, Guyatt, GH, Streiner, DL and King, DR (1997) Clinical Impact versus Factor Analysis for Quality of life Questionnaire construction. *Journal of Clinical Epidemiology. 50: 233-238.*

Kaplan, R, M and J, A, Ernest (1983) Do Category Rating Scales Produce Biased Preference Weights for a Health Index*? Medical Care, 21: 193-207.*

Kaplan, RM and JP, Anderson (1988) A General Health Policy Model: Update and Implications. *Health Serv Res. 23(2): 203-235.*

Kaplan, RM, CJ, Atkins, et al. (1984) "Validity of Quality of Well Being Scale as an Outcome Measure in Chronic Obstructive Pulmonary Disease." *J Chron Dis 37(2): 85-95.*

Kaplan, RM, D, Feeny, et al. (1993) "Methods for Assessing Relative Importance in Preference Based Outcome Measures." *Quality of Life Research 2: 467-475.*

Kaplan, RM, JW, Bush, et al. (1976) "Health Status: Types of Validity and the Index of Well-Being." *Health Services Research 11: 478-507.*

Katsivo, MN, LN, Muthami, et al. (1993) "Involvement of a Community in Schistosomiasis Control: *A Kenyan Experience." East African Medical Journal 70(8): 478-481.*

Katsivo, MN, LN, Muthami, et al. (1993) "Perception of a Schistosomiasis Control Project in Rural Kenya by the Beneficiaries." *East African Medical Journal 70(10): 613-616.*

Keasey, K (1984) Regret Theory and Information: A Note. *The Economic Journal, 94:645-648.*

KENQOL Group (1996) "Developing a Measure of HRQL in Kenya." *Health Economics and Financing Exchange .(No.9 Winter, 1996).*

Kerrigan, CL, Collins, ED, Kneeland, TS, Voigtlaender, D, Moncur, MM, Matheney, TH, Grove, MR and Tosteson, ANA (2000) Measuring Health State Preferences in Women with Breast Hypertrophy. *Plastic Reconstructive Surgery. 106: 280-8.*

Khanmann, D and A Tversky (1979) "Prospect Theory: An Analysis of decisions Under Risk." *Econometrica* 47: 263-91.

Kightlinger, LK, Seed, JR and Kightlinger, MB (1996) *Ascaris Lumbricoides* Aggregation in Relation to Child Growth Status, Delayed Cutaneous Hypersensitivity, and Plant Antihelminthic use in Madagascar. *Journal of Parasitology, 82(1):25-33.*

Kind, P (2003) Guidelines for Value sets in Economic and Non-economic Studies using EQ-5D. *In eds. Brooks R., Rabin R. and de Charro, F. The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective.*

Kirigia, JM (1994) The Economics of Schistosomiasis Interventions: A case Study of Mwea Irrigation Scheme in Kenya. PhD. Dissertation Department of Economics. York, University of York.

Kirigia, JM (1997) Economic Evaluation in Schistosomiasis: Using the Delphi Technique to Assess Effectiveness. *Acta Tropica 64: 175-190.*

Kirigia, JM (1998) "Cost-utility Analysis of Schistosomiasis Intervention Strategies in Kenya." *Environment and Development Economics 3: 319-346.*

Kirigia, JM (1998) Economic Evaluation in Schistosomiasis: Valuation of Health States Preferences- A Research Note. *Health Economics. 1998, 7(6): 551-558.*

Kirsch, J and McGuire, A (2000) Establishing Health State Valuations for Disease Specific States: An Example from Heart Disease. *Health Economics 9: 149-158.*

Kirstin, RJ (1999) Knowledge, Health and Progress Amongst and Akamba Population in Makueni District, Kenya. Ph.D. Thesis. South Bank University, UK.

Kloos, H, Ouma, JH et al. (1987) "Coping with intestinal illness among the Kamba in Machakos, Kenya, and aspects of schistosomiasis control." *Soc Sci Med 24(4): 383-94.*

Koutsoyiannis, A (1987) Modern Microeconomics. Second Edition. Chapter 2

Krabbe, P F, ML, Essink Bot, et al. (1996) "On the equivalence of collectively and individually collected responses: standard-gamble and time-trade off judgements of health states." *Med Decis Making, 16(2): 120-32.*

Krabbe, PF, ML Essink Bot, et al. (1997) "The comparability and reliability of five health-state valuation methods." *Soc Sci Med, 45(11): 1641-52.*

Krabbe, PFM. (1998) The Valuation of Health Outcomes: A Contribution to the QALY Approach. *Ph.D. Theses* Erasmus University. Rotterdam. *Chapter 2.*

Laidler, D (1981) Introduction to Microeconomics. Second E edition. Phillip Allan. *Chapter 4.*

Landgraf, J. M. and L. N. Abetz (1996) Measuring Health Outcomes in Paediatric Populations: Issues in Psychometrics and Application. *Quality of Life and Pharmacoeconomics in Clinical Trials. B. Spilker, Lippincot-raven Publishers.*

Lara-Munoz, MC, Sergio, P de L, and de la Fuente, JR (1995). The Development of an Instrument for the Measurement of the Quality of Life of Cancer Patients in Mexico. *Quality of Life Research. 4: 453.*

Lawless, JW, Latham, MC, Stephenson, LS, Kinoti, SN and Pertet, AM (1994) Iron Supplementation Improves Appetite and Growth in Anemic Kenyan Primary School Children. *Journal of Nutrition, 124:645-654.*

Lee, JE, Fos, PJ, Zuniga, MA, Kastl, PR and Sung, JH (2001) Assessing Health –Related Quality of Life in Cataract Patients: The Relationship between Utility and Health Related Quality of Life Measurement. *Quality of Life Research, 9: 1127-1135.*

Lenert, LA, Treadwell, JR and Schwartz, CE (1999) Associations Bewteen Health Status and Utilities: Implications for Policy. *Medical Care 37(5): 479-489.*

Leung, PP, Tannock, IF, Oza, AM, Poudziunas, A. and Dranitsaris, G (1990) Cost-Utility Analysis of Chemotherapy using Paclitaxel, Docetaxel, or Vinorelbine for Patients with Anthracycline-Resistant Breast Cancer. *Journal of Clinical Oncology, 17(10): 3082-3090.*

Llewellyn-Thomas, H, Sutherland, RN, Tibshirani, R, Ciampi, A, Till, JE and Boyd, NF (1982. The Measurement of Patients' Values in Medicine. *Medical Decision Making, 2(4): 449-462.*

Loewenstein, G, (1999) "Experimental Economics from the Vantage-Point of Behavioural Economics." *The Economic Journal 109(453): F25-F34.*

Loomes G and Sugden R (1987) Some Implications of a More General Form of Regret Theory. *Journal of Economic Theory 41:270-287.*

Loomes, G (1989).*Book Review* "Fishburn, PC (1988) Non-Linear Preference and Utility Theory". *Economica, Nov. p. 537.*

Loomes, G and R, Sugden (1982) "Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty." *Economic Journal 92: 805-24.*

Loomes, G, (1999) "Experimental Economics: Introduction." *The Economic Journal 109(453): F1-F4.*

Loomes, G, (1999) "Some Lessons from Past Experiments and Some Challenges for the Future." *The Economic Journal 109(453): F35-F45.*

Loomes, G, Starmer, C and Sugden, R (1992) Are Preferences Monotonic? Testing Some Predictions of Regret Theory. *Economica, 59:17-33.*

Loomes, G, Starmer, C and Sugden, R (1992) Are Preferences Monotonic? Testing Some Predictions of Regret Theory. *Economica, 59:17-33.*

Mackeigan, LD, O'Brien, B and Oh, PI (1999) Holistic Versus Composite Preferences for Lifetime Treatment Sequences for Type 2 Diabetes. *Medical Decision Making, 19: 113-121.*

Mahapatra, P, Salomon, JA and Nanda, L, (2002) Measuring Health State Values in Developing Countries—Results from a Community Survey in Andhra Pradesh. *In eds. Murray CJL, Salomon, JA, Mathers, CD and Lopez, AD (2002) Summary Measures of Population Health: Concepts, Ethics, measurement and Applications. WHO, Geneva.*

Manificat, S and A Dozard (1998) "Children's Quality of Life Assessment: Preliminary Results Obtained with the AUQUEI Questionnaire." *Quality of Life Newsletter, MAPI Research Institute.* (No.19.Jan-April.1998).

Manski, CF (1989) *Book Review* "Fishburn, PC.(1988) Non-Linear Preference and Utility Theory". *Journal of economic Literature, 17:1671-2.*

Marquis, P, Amerenco, G, Sapede, C, Jossserand, F, Jacquetin, B and Richard, F (1995c) Development and Validation of a Disease-Specific Quality of Life Questionnaire for Urinary Urge Incontinency. *Quality of Life Research, 4: 458.*

Marquis, P, Fayol, C and MaCarthy, C (1995b) Development and Validation of a Disease-Specific Short-Form Quality of Life Questionnaire for Intermittent Claudication. *Quality of Life Research, 4: 458.*

Marquis, P, Sapede, C, Bechade, D, Amouretti, M, Galmiche, JP, Poynard, T, Fiessinger, S, Girre, JP and Mathiex-Fortunet, H (1995a) Development and Psychometric Validation of a Disease-Specific Qualityt of Life Questionnaire in Gastro-oesophageal reflux. *Quality of Life Research, 4: 457.*

Matchar, DB, McCrory, DC and Bennett, CL (1997) Treatment Considerations for Persons with Metastatic Prostate Cancer: Survival Versus Out-of-Pocket Costs. *Urology, 49: 218-224.*

McDowell, I and Newell, C (1996) Measuring Health: A Guide to Rating Scales and Questionnaires, OUP.

McGuire, A, Henderson, J and Mooney, G (1988) The Economics of Health Care: An Introductory Text. Routledge. *Chapter 3.*

McLachlan, S, Pintilie, M and Tannock, IF (1999) Third line Chemotherapy in Patients with Metastatic Breast Cancer: An Evaluation of Quality of Life and Cost. *Breast Cancer Research and Treatment, 54:213-223.*

Medley, GF and DAP Bundy (1996) "Dynamic Modelling of Epidemiologic Patterns of Schistosomiasis Morbidity." *Am. J.Trop.Med. Hyg 55(5): 149-158.*

Mehrez, A and Gafni, A (1990) Evaluating Health Related Quality of Life: An Indifference Curve Interpretation for the Time Trade Off Technique. *Social Science and Medicine, 31(11): 1282-1283.*

Miller, AR, Cantor, SB, Peoples, GE, Pearlstone, DB and Skibber. JM (2000) Quality of Life and Cost Effectiveness Analysis of Therapy for Locally Recurrent Rectal Cancer. *Diseases of Colon and Rectum, 2000 43: 1695-1703.*

MOH (2002) Socio-economic and Health Indicators.
www.ministryofhealth.go.ke/indicators.html

Mohamed Ali, Q. E. Doehring Schwerdtfeger, et al. (1991) Ultrasonographical investigation of periportal fibrosis in children with Schistosoma mansoni infection: reversibility of morbidity seven months after treatment with praziquantel. *Am J Trop Med Hyg 44(4): 444-51.*

Mohide, EA, GW, Torrance, et al. (1988) "Measuring the wellbeing of family caregivers using the time trade-off technique." *J Clin Epidemiol 41(5): 475-82.*

Motulsky H, (1995) Intuitive Biostatistics, OUP (from www.graphpad.com/www/book/choose.htm)

Mugo, M and Fox-Rushby, JA (2003) Examining Content Validity and Conceptual Equivalence of Health State Valuation Techniques: Some Methodological Lessons. Paper Presented at IHEA 4th World Congress, San Francisco, USA.

Mukherjee, C, White, H and Wuyts M (1998) Econometrics and Data Analysis for Developing Countries. Rourledge. London and New York.

Mulkay, M, Ashmore, M and Pinch, T (1987) Measuring the Quality of Life: A Sociological Invention Concerning the Application of Economics to Health Care. *Sociology 21(4):541-564.*

Murray CJL and Lopez, AD (1994) Global Comparative Assessments in the Health Sector: Disease Burden, Expenditures and Intervention Packages. WHO.

Murray CJL and Lopez, AD (1996) The Incremental effect of Age-weighting on YLLs, YLD, and DALYs. *A Response. Bulletin of World Health Organization. 1996. 74(4): 445-446.*

Murray CJL and Lopez, AD (1997b) Regional Patterns of Disability-free Life Expectancy and Disability Adjusted Life Expectancy: Global Burden of Disease Study. *The Lancet. May 10, 1997, 349: 1347-52.*

Murray CJL, Salomon, JA, Mathers, CD and Lopez, AD (2002) Summary Measures of Population Health: Concepts, Ethics, measurement and Applications. WHO, Geneva.

Murray CJL, Tandon, A, Salomon, JA, Mathers, CD and Sadana, R (2002) New Approaches to Enhance Cross Population Comparability of Survey Results. In eds. *Murray CJL, Salomon, JA, Mathers, CD and Lopez, AD. Summary Measures of Population Health: Concepts, Ethics, measurement and Applications. WHO, Geneva.*

Murray, CJL (1996) Rethinking DALYs. *In eds. Murray CJL and Lopez, AD (1996) The Global Burden of Disease. Published by Harvard School of Public Health on behalf of World Bank and World Health Organization. Harvard University Press.*

Murray, CJL and Lopez, AD (1996) The Global Burden of Disease., the Harvard School of public Health on Behalf of World Bank and WHO. Harvard University Press.

Murray, CJL and Lopez, AD (1996) Evidence-Based health Policy: Lessons from the Global Burden of Disease Study. Science, 1996. 274: 740-743.

Muthami, LN, MN, Katsivo, et al. (1995) "Pattern of Schistosoma mansoni infection after intervention in Mwea irrigation scheme in Kenya." *East Afr Med J 72(2): 94-100.*

Nookes, C and Bundy, DAP (1994) Does Helminth Infection Affect Mental Processing and Educational Achievement? *Parasitology Today, 10(1):14-18.*

Nord, E (1992a) An Alternative to QALYs: The Saved Young Life Equivalent (SAVE*). British Medical Journal 305: 875-7.*

Nord, E (1992b) Methods for Quality Adjustment of Life Years." *Social Science and Medicine 34(5): 559-569.*

Nord, E (1994) "The QALY - A Measure of Social Value Rather than Individual Utility?" *Health Economics 3: 89-93.*

Nord, E (1995) "The Person Trade-Off Approach to Valuing Health Care Programs." *Medical Decision Making 15: 201-208.*

Nord, E, JL, Pinto, et al. (1999) "Incorporating societal concerns for fairness in numerical valuations of health programmes." *Health Econ 8(1): 25-39.*

O'Brien, BJ (1990) Economic Approaches to the Problems of Acceptable Clinical Risks: The Case of Prescription Drugs and Chronic Rheumatic Disease. *Ph.D. Theses.* Brunel University, Department of Economics. *Chapters 1 and 2*

Oldridge, N, Furlong, W, Feeny, D, Torrance, G, Guyatt, G, Crowe, J and Jones, N (1993) Economic Evaluation of Cardiac Rehabilitation Soon After Acute Myocardial Infarction. *The American Journal of Cardiology, 72: 154-161.*

Pal, DK (1996) "Quality of Life Assessment in Children: A review of Conceptual and Methodological issues in multi-dimensional Health Status measures." *J Epidemiol Community Health 50: 391-396.*

Parker M and Fox-Rushby JA (1995) International Comparisons in Health-Related Quality of Life: Acquiescence in Academia. *In Eds Shumaker, SA and Berzon, R (1995) The*

International Assessment of Health Related Quality of Life: Theory, Translation, Measurement and Analysis. Rapid Communications of Oxford.

Parker, M (1992) "Reassessing Disability: The Impact of Schistosomal Infection on Daily Activities Among Women in Gezira Province, Sudan." *Social Science and medicine, 35*(7): *877-890.*

Parker, R and Hopwood, P (2000) Literature Review: Quality of life of Black and Ethnic minority Groups with Cancer. Cancer Research UK Psychological Medicine Group, University of Manchester.

Patrick, D, Bush, JW, and Chen, MM (1973) "Methods for Measuring Levels of Well-Being for a Health Status Index." *Health Services Research, 228-245.*

Patrick, D, Starks, HE, Cain, KC, Uhlmann, RF and Pearlman, RA (1994) "Measuring Preferences for Health States Worse than Death." *Medical Decision Making, 14: 9-18.*

Patrick, DEL and P, Erickson (1993) Health Status and Health Policy: Quality of life in Health Care Evaluation and Resource Allocation, *OUP.*

Patrick, DL and RA Deyo (1989) "Generic and Disease-specific Measures in Assessing Health Status and Quality of Life." *Medical Care 27(3): S217-S232.*

Perez, DJ, McGee, R, Campbell, AV, Christensen, EA and Williams, S (1997) A Comparison of Time Trade-Off and Quality of Life Measures in Patients with Advanced Cancer. *Quality of Life Research, 6(2): 133-138.*

Pett, MA (1997) Nonparametric Statistics for Health Care Research: Statistics for Small Samples and Unusual Distributions. SAGE publications .

Pinto, JP (1997) "Is the Person Trade-Off a valid Method for Allocation Health Care Resources?" *Health Economics 6: 71-81.*

Rabin, R, Herdman, M, Fox-Rushby, J and Badia, X (2003) Exploring the Results of Translating the EQ-5D in 11 European languages. *In eds. Brooks R., Rabin R. and de Charro, F. (2003) The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective.*

Raj, SM, Sein, KT, Anuar, AK et al. (1996) Effect of Intestinal Helminthiasis on Intestinal Permeability of Early Primary School Children. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 90: 666-669.*

Rawles, J (1989) Castigating QALYs. *Journal of Medical Ethics. 15:143-147.*

Read, JL, Quinn, RJ et al. (1984) "Preferences for Health Outcomes. Comparison of Assessment Methods." *Medical Decision Making 4(3): 315-29.*

Read, JL, Quinn, RJ et al. (1987) "Measuring Overall Health: An Evaluation of Three Important Approaches." J Chron Dis 40(suppl 1): 7S-21S.

Revicki, DA and Kaplan, RM (1993) Relationship Between Pychometric and Utility Based Approaches to the Measurement of Health Related Quality of Life. *Quality of Life Research, 2: 477-487.*

Revicki, DA Leidy, NK, Brennan-Diemer, F, Sorensen, S and Togias, A. (1998) Intergrating Patient Prefernces into Health Outcome Assessment* The Multiattribute Asthma Symptom Utility Index. *Chest, 114:998-1007.*

Revicki, DA and Wood, M (1998) Patient-Assigned Health State Utilities for Depression-Related Outcomes: Differences by Depression Severity and Antidepressant Medications. *Journal of Affective Disorders, 48: 25-36.*

Richardson, J (1994) Cost Utility Analysis: What Should be Measured? *Social Science and medicine, 39(1):7-21.*

Richardson, J and E Nord (1997) "The Importance of Perspective in the Measurement of Quality Adjusted Life Years." *Medical Decision Making 17: 33-41.*

Robinson, A, P, Dolan, et al. (1997) "Valuing health status using VAS and TTO: what lies behind the numbers?" *Soc Sci Med 45(8): 1289-97.*

Robinson, A., Loomes, G. and Jones-Lee, M. Visual Analogue Scales, Standard Gambles, and Relative Risk Aversion. Medical decision Making 2001, 21:17-27.

Rosenbaum, PL and S, Saigal (1996) Measuring Health -related Quality of Life in Paediatric Populations: Conceptual Issues. *Quality of Life and Pharmacoeconomics in Clinical Trials. B, Spilker, Lippincot-raven Publishers.*

Rosser, R and Kind, P (1978) A Scale of Valuations of States of Illness: Is there a Social Consensus? *International Journal of Epidemiology, 7(4):347-358.*

Rutten-van Molken, MPMH, Bakker, CH, van Doorslaer, EKA, and van der Linden, S (1995) Methodological Issues of Patient Utility Measurement: Experience from Two Clinical Trials. *Medical Care, 33(9): 922-937.*

Sackett, DL and Torrance, GW (1978) The Utility of Different Health states as perceived by the General Public. *Journal of Chronic Diseases, 31:697-704.*

Saconato, H and Atallah, A (1999) Interventions for Treating Schistosomiasis Mansoni (Cochrane Review). *In The Cochrane Library, Issue3. Oxford: Update software.*

Sadana R (2002) Measurement of Variation in Health State Valuations in Phnom Penh, Cambodia. In eds. *Murray CJL, Salomon, JA, Mathers, CD and Lopez, AD. Summary Measures of Population Health: Concepts, Ethics, measurement and Applications. WHO, Geneva.*

Saigal, S, BL, Stoskopf, et al. (1999) "Differences in Preferences for Neonatal Outcomes Among health care Professionals, Parents and Adolescents." JAMA 281(21): 1991-1997.

Salomon JA and Murray CJL (2002) A Conceptual Framework for Understanding Adaptation, Coping and Adjustment in Health State Valuations. In eds. *Murray CJL, Salomon, JA, Mathers, CD and Lopez, AD. Summary Measures of Population Health: Concepts, Ethics, measurement and Applications. WHO, Geneva.*

Salomon JA and Murray, CJL (2002) Estimating Health State Valuations using a Multiple-Method Protocol. In eds. *Murray CJL, Salomon, JA, Mathers, CD and Lopez, AD 2002. Summary Measures of Population Health: Concepts, Ethics, measurement and Applications. WHO, Geneva.*

Sauraez-Almazor, ME and Corner-Spady, B (2001) Rating of Arthritis Health States by Patients, Physicians, and the General Public. Implications for Cost Utility Analysis. *Journal of Rheumatology. 28; 648-56.*

Sen, A (1982) Choice, Welfare and Measurement. Basil Blackwell. *Chapters 2, 3, and 4.*

Sherbourne, CD, Keeler, E, Unutzer, J, Lenert, L and Wells, KB (1999) Relationship Between Age and Patients' Current Health State Preferences. *Gerontologist. 39(3): 271-278.*

Shumaker, SA and Berzon, R (1995) The International Assessement of Health-Related Qualityof Life: Theory, Translation, Measurement and Analysis. Rapid Communications.

Shumaker, SA and Naughton, MJ (1995) The International Assessement of Health-related Quality of Life: A Theoretical Perspective. In eds. Shumaker, SA and Berzon, R, The international Assessement of Health-Related Quality of Life: Theory, Translation, Measurement and Analysis. Rapid Communications.

Shweder, RA and Bourne EJ. Does the Concept of the Person Vary Cross-Culturally? In: Shweder RA, Levine RA eds. Cultural Theory: Essays onMind, Self and Emotion. Cambridge university Press, 1984.

Skevington SM. Advancing Cross-cultural Research on Quality of Life: Observations Dawn for the WHOQOL Development. Quality of Life Research, 2002, 11:135-144.

Skevington, S, MacArthur and Somerset, M (1995) Developing the WHOQOL in Britain: A study of Contemporary Beliefs about Quality of Life. *Quality of Life Research. 1995 4: 485.*

Sloan, FA, (1996) Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies. Cambridge University Press.

Smith, RD (1996) "Is Regret Theory an Alternative Basis for Estimating the Value of Health Care Interventions?" *Health Policy, 37: 105-115.*

Souchek, J, Stacks, JR, Brody, B, Ashton, CM, Giesler, B, Byrne, MM, Cook, K, Geraci, JM and Wray, NP (2000) A Trial for Comparing Methods for Eliciting Treatment Preferences from Men with Advanced Prostate Cancer: Results from the Initial Visit. *Medical Care, 2000. 38(10): 1040-1050.*

Spencer, A. The Implications of linking Questions within the SG and TTO methods. Health Economics, (in press) 2004.

Starmer, C (1999) "Experimental Economics: Hard Science or Wasteful Tinkering?" *The Economic Journal 109(453): F5-F15.*

StataCorp. (2001) Stata Statistical Software: Release 7.0. College Station, TX: Stata Corporation

Stenner, PHD, Cooper, D and Skevington SM. Putting the Q into Quality of Life; the Identification of Subjective Constructions of Health related Quality of Life using the Q methodology. Social Science and Medicine, 2003, 57:2161-2172.

Stephenson, L (1989) Africa. In eds. Crompton, DWT, Nesheim, MC and Pawlowski, ZS (1989) Ascaris and its Prevention and Control. Taylor and Francis, London.

Stephenson, L (1993) The Impact of Schistosomiasis on Human Nutrition. *Parasitology 1993 107:S107-S123*

Stephenson, L, Latham, MC, Adams, EJ, Kinoti, SN And Pertet, A (1994) Physical Fitness, Growth and Appetite of Kenyan School Boys with Hookworm, Trichuris Trichuria and Ascaris lumbricoides infections Are improved Four months after a Single Dose of Albendazole. *Journal of Nutrition. 123: 1036-1046.*

Stephenson, L, Latham, MC, Kinoti, SN, Kurz, KM and Brigham, H (1990) Improvements in Physical Fitness of Kenyan school Boys Infected with Hookworm, trichuris Trichuria and Ascaris Lumbricoides Following a Single Dose of Albendazole. *Transactions of the Royal Society of Tropical Medicine and Hygiene 84: 277-282.*

Stephenson, L. (1993) "The impact of schistosomiasis on human nutrition." *Parasitology 107 Suppl: S107-23.*

Stephenson, LS and C, Holland (1987) The impact of helminth infections on human nutrition schistosomes and soil-transmitted helminths. London Taylor & Francis.

Stiggelbout, A, Kiebert, GM, Kievit, J, Leer, JWH, Stoter, G and DE Haes, JCJM (1994) "Utility Assessment in Cancer Patients: Adjustment of Time Trade Off Scores for the Utility of Life Years and Comparison with Standard Gamble Scores." *Medical Decision Making, 14: 82-90.*

Stolk, EA, Busschbach, JJV, Caffa, M, Meuleman, EJH, Rutten, FFH (2000) Cost Utility Analysis of Sildenafil Compared with Papaverine-Phentolamine injections. *British Medical Journal. 320: 1-6.*

Stoltzfus, RJ, Albonico, A, HM, Tielsch, JM, et al. (1997a) School Based Deworming Program yields Small improvement in Growth of Zanzibari School Children After One Year. *Journal of Nutrition, 127:2187-2193.*

Stoltzfus, RJ, Albonico, M, Chwaya, HM, et al. (1996) Hemoquant Determination of Hookworm-related Blood Loss and its Role in Iron Deficiency in African Children. *American Journal of tropical medicine and Hygiene, 55(4):399-404.*

Stoltzfus, RJ, Chwaya, HM, Tielsch, JM, et al. (1997b) Epidemiology of Iron Deficiency Anemia in Zanzibari school Children: the Importance of Hookworms. *American Journal of Clinical Nutrition, 65:153-9.*

Streiner, GL and DR Norman (1995) Health Measurement Scales: A Practical Guide to their Development and Use. Oxford. OUP.

Sugden, R (1989) *Book Review* "Fishburn, PC (1988) Non-Linear Preference and Utility Theory". *The Economic Journal, p. 1191-2.*

Sukwa, YT, Bulsara, MK and Wurapa, FK (1986) The Relationship Between Morbidity and Intensity of *Schistosomiasis Mansoni* Infection in a rural Zambian Community. *International Journal of Epidemiology 15:248-251.*

Swan, JS, Fryback, DG, Lawrence, WF, Katz, DA, Heisey, DM, Hagenauer, ME, Ellzer, PM and Jacobson, BK (1997) MR and Conventional Angiography: Work in Progress Toward Assessing Utility in Radiology. *Academic Radiology. 4: 475-482.*

Swan, JS, Fryback, DG, Lawrence, WF, Sainfort, F, Hagenauer, ME and Heisey, DM (2000) A Time-Tradeoff Method for Cost-Effectiveness Models Applied to Radiology. *Medical Decision Making. 20: 79-88.*

Swenson, MD (1992) Scarcity in the Intensive Care Unit: Principles of Justice for Rationing ICU beds. *The American Journal of Medicine, 92:551-555.*

Tanner, M (1989) "Evaluation of public health impact of schistosomiasis." *Trop Med Parasitol 40(2): 143-8.*

Tan-Torres, TL (1991) Comparison of Different Methods of Eliciting Utilities for Outcome States in Leprosy. Unpublished report, Geneva, TDR / WHO.

Tanumihardjo, SA, Permaesih, D, et al. (1996) Vitamin A Status of Indonesian Children with *Ascaris lumbricoides* after dosing with Vitamin A Supplements and Albendazole. *Journal of Nutrition, 126:451-457.*

The general Model: Biological–Psychological–Social.
http://www.qeeg.com/biopsychosocial.html 01/02/2004.

Tolley, G, Kenkel, D and Fabian, R (1994) Valuing Health for Policy: An Economic Approach. *The University of Chicago Press.*

Torrance, G, D, Feeny, et al. (1996) "Multiattribute Utility Function for a Comprehensive Health Status Classification System: Health Utilities Index Mark 2." *Medical Care 34(7): 702-722.*

Torrance, GW, Feeny, D and Furlong, W (2001) Visual analog scales: do they have a role in the measurement of preferences for health states? *Medical Decision making 21(4): 329-34*

Torrance, GW (1976) Social Preferences for Health States: An Empirical Evaluation of Three Measurement Techniques. *Socio-Economic Planning Sciences, 10:129-136.*

Torrance, GW (1986) "Measurement of Health State Utilities for Economic Appraisal." *Journal of Health Economics 5: 1-30.*

Torrance, GW (1987) Utility Approach to Measuring Health Related Quality of Life. *Journal of Chronic Diseases, 40(6):593-600.*

Torrance, GW (1996) Designing and Conducting Cost Utility Analyses. *In eds. Spilker, B, Quality of Life and Pharmacoeconomics in Clinical Trials. Second edition. Chapter 114.*

Torrance, GW, Thomas, WH and Sackett, DL (1972) A Utility Maximisation Model for Evaluation of Health care Programs. *Health Services Research, 118-133.*

Truchon, M. Determinants of Chronic Disability Related to Low Back Pain: Towards an Integrative Biopsychosocial Model. Disability and Rehabilitation, 2001, 23(17): 758-476.

Tsevat, J, NV, Dawson, et al (1998) "Health values of hospitalized patients 80 years or older. HELP Investigators. Hospitalized Elderly Longitudinal Project [see comments]." *JAMA 279(5): 371-5.*

Ubel, PA, Richardson, J and Menzel, P (2000) Societal Value, the Person Trade-Off and the Dilemma of Whose Values to Measure for Cost Effectiveness Analysis. *Health Economics, 9: 127-136.*

UN OCHA IRIN. 2002 East Africa: Bilharzia Health Hazard Highlighted. *www.Reliefweb.int/w/rwb.nsf*

Unic, I, PF, Stalmeier, et al. (1998) "Assessment of the time-trade off values for prophylactic mastectomy of women with a suspected genetic predisposition to breast cancer." *Medical Decision Making 18(3): 268-77.*

Ustun, B, Rehm, J and Chatterji, S (2002) Are Disability Weight Universal? Ranking of Disabling Effects of Different Health Conditions in 14 Countries by Different Informants. . In eds. *Murray CJL, Salomon, JA, Mathers, CD and Lopez, AD 2002. Summary Measures of Population Health: Concepts, Ethics, measurement and Applications. WHO, Geneva.*

Varian, HR (1990).Intermediate Microeconomics: A Modern Approach. Second Edition *Chapters 3, 4, 7, and 12.*

Von Neumann, J and Morgenstern, O (1953) Theory of Games and Economic Behavior. Princeton University Press. *Chapter 1.*

Walker, D and Fox-Rushby JA (2000) Economic Evaluation of Communicable Disease Interventions in developing Countries: A Critical Review of the Published Literature. *Health Economics, 9(8): 681-698.*

Warren, KS, Bundy, DAP et al (1993). Helminth infections. *In eds Jamison, DT et. al. (1993) Disease Control Priorities in Developing Countries, 1993. Oxford. OUP.: 131-160.*

Watkins, WE, Jose, RC and Pollitt, E (1996) The Effects of Deworming on Indicators of School Performance in Guatemala. *Transactions of the Royal Society of Tropical Medicine and Hygiene. 90: 156-161.*

What Does the Cronbach's alpha mean? http://www.ats.ucla.edu/stat/spss/faq/alpha.html.

WHO (1985) The Control of Schistosomiasis. Geneva, WHO.

WHO (1993) The Control of Schistosomiasis. Geneva, WHO.

WHO (1993) The Control of Schistosomiasis. Geneva, WHO.

WHO (1993) World Development Report 1993: Investing in Health. *OUP.*

WHO (1998) Life in the 21st Centuary: A Vision for All. The World Health Report. Geneva, WHO.

WHO/CTD (1998) Intestinal Parasites Control., http://www.who.int/ctd/html.

WHO/CTD (1999) Schistosomiasis Control, http://www.who.int/ctd/html/schisto.html.

Wikipedia, the free encyclopendia. Biopsychosocial Model.
http://en.wikipedi.or/wiki/biopsychosocial_model. 01/02/ 2004

Wilkin, D, Hallam, L and Doggett, M (1992) Measures of Need and Outcome for Primary Health Care. *OUP.*

Williams A (1996) QALYs and Ethics: A Health economist's Perspective. *Social Science and Medicine 43(12):1795-1804.*

Williams, A (1995) The Role of the EuroQoL Instrument in QALY Calculations. Discussion Paper 130. Center for Health Economics, York Health Economics Consortium, NHS Center for Reviews & Dissemination. The University of York.

Wolfson, AD, Sinclair, AJ, Bombardier, C and McGeer, A (1982) Preference Measurements for Functional status in Stroke Patients: Interrater and Intertechnique

Comparisons. In Eds. Kane, RL and Kane RA (1982). Values and Long-Term Care. Lexington, Mass. 1982:191-214.

World Bank Group. (2003) Country-Competitiveness Indicators: Kenya. *http://wbln0018.worldbank.org/psd/compete.nsf*.

World Bank, (2003) World Development Indicators Database. www,worldbank.org/data/

Wright, JG. (1996) Quality of Life in Orthopedics. *Quality of Life and Pharmacoeconomics in Clinical Trials. B. Spilker, Lippincot-raven Publishers.*

Wu, AW, Cagney, KA, Fink, NE, Meyer, KB, Herbert, RJ, Jenckes, MW, Sadler, JH and Powe, NR (1995) Development of a Quality of Life Measure for Renal Dialysis patients. *Quality of Life Research. 4: 505.*

Wu, G (1996) The Strengths and Limitations of Expected Utility Theory. *Medical Decision Making, 16:9-10.*

Yamada, S., Green G., Bauman, K. and Maskarinec, G. A Biopsychosocial Approach to Finding Common Ground in the Clinical Encounter. Academic Medicine, 2000, 75(6):643-648.

Yee, GC (1997) Cost-Utility Analysis of Taxane Therapy. *American Journal of Health System Pharmacy. 54(suppl. 2):S11-15.*

Zug, KA, Littenberg, B, Baughman, RD, Kneeland, T, Nease, RF, Sumner, W, O'Connor, GT, Jones, R, Morrison, E and Cimis, R (1995) Assessing the Preferences of Patients with Psoriasis: A Quantitative, Utility Approach. *Archives of Dermatology, 131: 561-568.*

APPENDIX 1.1 DISTRIBUTION OF SCHISTOSOMIASIS IN KENYA

Table A1.1: Outpatient Morbidity from Schistosomiasis

PROVINCE	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
NAIROBI	0%	0%	1%	0%	0%	0%	0%	4%	1%	3%	3%	0%
CENTRAL	**6%**	**6%**	**21%**	**8%**	**33%**	**7%**	**7%**	**20%**	**6%**	**11%**	**10%**	**9%**
COAST	45%	65%	40%	65%	37%	48%	51%	47%	53%	44%	56%	41%
EASTERN	9%	7%	18%	10%	9%	16%	15%	3%	9%	11%	2%	15%
N/EASTERN	2%	1%	2%	3%	3%	4%	5%	4%	6%	7%	3%	6%
NYANZA	18%	10%	12%	12%	5%	11%	10%	6%	14%	9%	16%	19%
R/VALLEY	9%	5%	6%	2%	12%	13%	4%	8%	6%	9%	10%	6%
WESTERN	10%	2%	1%	0%	1%	0%	7%	8%	5%	6%	0%	4%
TOTAL SCHISTOSOMIASIS	113,707	67,602	105,439	59,451	93,164	81,825	53,928	57,780	62,289	37,363	24,461	28,380
TOTAL ALL CASES	28,745,182	1,516,356	19,095,664	17,373,882	17,568,249	21,797,865	18,744,323	20,995,984	22,556,899	19,181,835	12,427,180	15,705,823
% OF TOTAL CASES	**0.4**	**0.43**	**0.55**	**0.34**	**0.53**	**0.4**	**0.28**	**0.38**	**0.28**	**0.19**	**0.2**	**0.18**

Source: Statistical Abstracts, 1991, 1994, 1995 and MOH Health Information System Report 1995

Table A1.2: Distribution of *Schistosomiasis Mansoni* in Central Province, by District (outpatient Morbidity Statistics): 1995

District	cases	% contribution schistosomiasis	% contribution province
Kiambu	482	19.0	0.015
Kirinyaga	**1072**	**42.2**	**0.034**
Murang'a	144	5.7	0.005
Nyandarua	20	0.8	0.001
Nyeri	822	32.4	0.026
Total (schistosomiasis)	2540		
Total cases (Province)	3,143,366		
% of Schistosomiasis	**0.08**		

APPENDIX 3.1: CRITERIA FOR JUDGING MEASUREMENT AND VALUATION TOOLS

It is essential to demonstrate that measurement instruments measure what they aim to measure with least amount of error. Ensuring equivalence of instruments, validity, reliability and practicality enhances the credibility of the instruments as well as the data they generate. These aspects are presented below.

3.1 Validity

Validation of a technique is particularly important in situations where there is no agreement about what is being measured, such as quality of life (Streiner and Norman, 1995). Validity refers to the ability of a measurement instrument to measure what it purports to measure, namely the value attached to a health-state (Buckingham et al. 1996). Validity can be established through content validity[73], face validity, criterion validity assessed through concurrent and predictive validity, and construct validity assessed through convergent and discriminant validity (Carmines and Zeller, 1979: Streiner and Norman, 1995). "Strictly speaking one does not assess the validity of a measure but the use it is being put to" (Carmines and Zeller, 1979 p.12). Following is a brief discussion of each validity type along with assessment methods.

3.1.1 Content Validity

Content validity depends on the extent to which an empirical measurement reflects a specific domain of content (Carmines and Zeller, 1979). Hence, in health state measurement and valuation it refers to whether the description of the health states is relevant and comprehensive in terms of content of domains and whether values obtained really represent individuals preferences for those health states. Content validity depends on the definition of the concept being measured, health. Since, there is no single accepted definition of what health encompasses, content validation in health related quality of life work is severely limited, unless a clear concept exists.

Steps in assessing content validity include specifying the full domain of content that is relevant to the particular measurement situation, sampling specific aspects of the domains to reflect the situation and putting them in a form that is testable. As noted in Herdman et al. (1998), Guyatt (1995) and Bullinger et al. (1993) content and face validity require qualitative data and depends on judgement by experts and validity by assumption (Streiner and Norman, 1995). "*Inevitably content validity rests mainly on appeals of reason regarding the adequacy with which important content has been sampled ... ... and ...cast in the form of test items" (Nunnally, 1978: 93 in* Carmines and Zeller, 1979: 22). Hence, one can not say the specific extent to which a measure should be considered content valid. However, this should be judged by examining the steps followed in ensuring that relevant domains and item content considered important by those the tools are targeted at, are included.

[73] In HRQL, there is no agreed definition of health and content therefore depends on judgements by experts and validity by assumption (Streiner and Norman, 1995).

3.1.2 Criterion Validity

Criterion related validity "is at issue when the purpose is to use an instrument to estimate some form of behavior that is external to the measuring instrument itself, the later being referred to as the criterion" (Carmines and Zeller, 1979:17). Criterion validity depends on the extent of correspondence between the test and criterion (Carmines and Zeller, 1979). Concurrent validity correlates a new measure with the criterion measure both given at the same point in time. This can for example be tested by applying the instruments concurrently and establishing the correlation of their results (Torrance, 1972) although there is no way of telling that the two instruments are measuring health. Establishing concurrent validity is useful when a new measure is constructed that claims to be better than those existing are (Fink, 1993). Predictive validity correlates a new measure with a criterion that is not available until some time in the future.

The extent of criterion validity is assessed by the degree of correspondence between the test and the criterion is usually estimated by the size of their correlation, which ranges between 0 to 1. The higher the correlation, the more valid is this test for this particular criterion (Carmines and Zeller, 1979) although Bowling (1997) notes that a correlation of 0.5 would be considered acceptable. Selection of the criterion variables should be based on some behavioral theory dictating how the variables relate. Criterion validation in social sciences is limited because there do not exist relevant criterion variables. As Krabbe et al. (1997) and Froberg and Kane, (1989b) argue, criterion and content validity can not be evaluated in HRQL instruments as they are part of an on-going process.

3.1.3 Construct Validity

In the absence of a gold standard, the most rigorous approach to establishing validity is testing construct validity (Dolan et al. 1996a). A construct is a theoretically derived notion of what the method is intended to measure. Construct validity is concerned with the extent to which a measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts that are being measured. It must be conceived of within a theoretical context and performance assessed in accordance with theoretical expectations (Carmines and Zeller, 1979).

Construct validity examines the extent to which an instrument is related to other measures in the way we expect (convergent validity) and whether the instrument is influenced by other factors (discriminant validity) (Buckingham et al. 1996). Construct validation is a process of hypothesis testing and it assesses both theory and the method simultaneously. Three steps in assessing construct validity include specifying the theoretical relationships between constructs, examining the empirical relationships between the measures of the concepts and interpreting the empirical evidence in terms of how it clarifies the construct validity of the measure (Carmines and Zeller, 1979). Careful interpretation of findings is required as the problem could be with the validity of the method or the theory (Bowling, 1997). Carmines and Zeller (1979) provide four possible interpretations if evidence relevant to construct validity is negative. One, that the measure lacks construct validity for this

particular theoretical construct, and hence should not be used as an empirical manifestation of that concept in future researches. Two, that the theoretical framework used to generate the empirical predictions is incorrect, thereby casting doubt on the underlying theoretical perspective. Three, that the method or statistical procedure used to test the theoretically derived hypothesis is faulty or inappropriate but the constructs are properly stated and a relationship exists between them. Finally, there is lack of construct validity or the unreliability of some other variables in the analysis e.g. measuring one of the constructs wrongly.

Two approaches that have been used in construct validation include examining the extent to which the results of different scaling methods converge and examining the degree to which predicted relationships between preferences and other variables are empirically supported (Froberg and Kane (1989b), as in Mohide et al. (1988), Dolan et al. (1996a), Torrance (1972), Badia et al. (1999), Krabbe et al. (1997) and Patrick et al. (1973).

Statistical procedures such as analysis of variance, comparisons of mean values and correlation analysis have been used to assess validity. In terms of statistical tests for validity, correlation coefficients such as product-moment Pearson correlation, Spearman-rank correlation and interclass correlation (Krabbe et al. 1997) have been used especially in tests of construct validity. However, there is no one experimental design or statistic that is common to construct validation studies (Streiner and Norman, 1995). As correlation can be positive or negative the values for correlation coefficients vary between 0 and 1 for convergent validity and –1 and 0 for discriminant validity. The closer the coefficients are to 0 the less the validity being established.

Construct validation is an on-going process, of learning more about the construct, making new predictions and then testing them (Streiner and Norman, 1995). As Carmines and Zeller (1979. P 24) note, "construct validity is not established by confirming a single prediction on different occasions or confirming many predictions on in a single study" It requires a pattern of consistent findings involving different researchers using different theoretical structures across a number of different studies.

3.2 Reliability

Reliability of a measure reflects the amount of error, both random and systematic, inherent in the measurement. It establishes stability of values within and between raters as well as over time when there is no evidence of change. Reliability of a method can be tested in several forms (Dolan et al. 1996a: Carmines and Zeller, 1979) and these are test-retest, inter-rater, intra-rater, alternative form, split halves and internal consistency. In the next sub-sections these forms of reliability are described followed by a presentation of analytical approaches for assessing reliability.

3.2.1 Test-retest reliability

Test retest reliability is the commonest form of reliability used to evaluate measurement and valuation tools. It assesses the stability of values over short periods of time, by repeating the test on a representative sub-sample of the study population. The time between test and retest should be selected such that things will not have changed and subjects do not remember their previous answers (Streiner and Norman, 1995). It is assumed that responses to the test will correlate across time because they reflect the same true variable. Hence reliability is equal to the correlation between the scores on the same test obtained at two points in time. The higher the correlation the better the reliability.

A conceptual difficulty in establishing test-retest reliability is in determining how much time to allow between the two tests. If too much time is allowed, external events might influence the response for the second test and if too little time lapses, respondents might remember their responses form first test and simply repeat them (Fink, 1993). A retest interval of two to 14 days is usual (Streiner and Norman, 1995) although Carmines and Zeller, (1979. P. 40) suggest that two weeks to 1 month is advisable to complete both tests. However, this duration can vary depending on the study circumstances and hence there is no consensus on the exact duration between tests for establishing test-retest reliability.

Test-retest reliability suffers some limitations. First, there is the possibility of change in the underlying concept itself resulting in low correlations. Secondly, reactivity, whereby the process of measuring a phenomenon induces change in the phenomena itself can result in poor correlation and thirdly, memory effects whereby the memory of the responses during the first interview influences responses in the second (Carmines and Zeller, 1979). These influences require consideration in interpreting and reporting reliability of a tool.

3.2.2 Inter-rater reliability

This refers to the degree of agreement between different observers on separate occasions (Brooks, 1995). Therefore, examining the extent to which, results obtained by two or more interviewers agree for similar or the same populations on different occasions assess inter-rater reliability. Implicit in this definition is that there is no time interval between the different observations (Brooks, 1995). Therefore to assess this form of reliability would ideally entail administering the interviews more than once on the same respondent by different interviewers on the same day at different occasions. However, there are no observers involved in using HRQL scales (Streiner and Norman, 1995). Hence, these requirements present logistical difficulties in assessing inter-rater reliability in HRQL studies, and hence not commonly assessed. However, a less ideal alternative would be to compare values from different respondents by different interviewers for each test administration, i.e. first and second tests separately, rather than comparing overtime which would be same as test-retest reliability.

3.2.3 Intra-rater reliability

This refers to the degree of agreement of observations made by the same observer on separate occasions (Brooks, 1995). It tests for individual respondent's consistency in responses (Froberg and Kane, 1989b: Dolan et al. 1996a; Fink, 1993), since it measures variation which occurs within an observer as a result of multiple exposures to the same stimulus (Streiner and Norman, 1995). Implicit in this definition is that there is no time interval between the different observations (Brooks, 1995). Logistical difficulties similar to those in inter-rater reliability are also present in assessing this form of reliability. However, a less ideal alternative would be to assess the extent to which responses from respondents interviewed by a particular interviewer agree, within the first and second test separately. This would reveal whether each interviewer applied different standards on different occasions (Streiner and Norman, 1995).

3.2.4 Alternative forms Method

Alternative form method requires two testing situations with the same people (Carmines and Zeller, 1979. P. 37) except that a different test (alternative form) is given on each occasion. The two tests are intended to measure that same thing and correlations between the two forms are an estimate of reliability. This test is intended to take care of memory effects but is limited in that it is difficult to develop one test, much less two forms. Also the fact that it is administered in two points in time means that we cannot distinguish true change from unreliability of the measure.

3.2.5 Split Halves Method

This test is conducted on one occasion. The total set of items is split into half and their scores correlated to obtain an estimate of reliability. The halves are considered approximations of alternative forms. There are different ways of splitting the items and each split is likely to give different reliability result. Hence there is an element of indeterminacy of reliability using this technique.

3.2.6 Internal Consistency

Internal consistency also referred to as homogeneity (Fink, 1993) is the extent to which the items comprising a scale measure the same thing, such as characteristic, skill of quality (Fink, 1993). It does not require retest or splitting the items and is administered once. The most popular of these reliability estimates is given by Cronbach's alpha (α). It varies between 0-1. α depends on the average inter correlation among all of the items. The correlation is calculated to determine the extent of homogeneity (Fink, 1993) and therefore a low coefficient alpha (α) indicates that the item does not come from the same conceptual domain. Carmines and Zeller (1979) state that reliabilities should not be below 0.8 for widely used scales.

3.2.7 Analytical Approaches for Assessing Reliability

The analytical approach to calculating reliability is based on the statistical technique of analysis of variance (ANOVA) (Streiner and Norman, 1995). The statistics used include the inter-class correlation coefficient computed as subject variability divided by the sum of subject variability and measurement error. Pearson's product-moment correlation coefficient is used for test re-test reliability where the data is continuos. It is based on regression analysis and is a measure of the extent to which the relationship between two variables can be explained by a straight (regression) line (Streiner and Norman, 1995). The kappa coefficient or test of concordance is suitable for data that is not a continuum (categorical) but where there is absence of presence of an attribute. It is useful for evaluation inter-rater and intra-rater reliability (Bowling, 2001.p20) This measure calculates simple agreement, the proportion of responses in which the two observations agreed (Streiner and Norman, 1995). Spearman's Rho and Kendall's coefficient of concordance is suitable with ranked (ordinal) sets of data. Statistical tests for internal consistency, split halves and alternative forms include inter-items correlations and Cronbach alpha (Bowling, 2001). There are currently suggestions to assess reliability using confidence intervals to assess the size of the difference between scores (Bland and Altman, 1986 and Ruta et al. 1994a in Bowling, 2001).

3.3 Levels for acceptable validity and reliability coefficients

There are no set criteria for acceptable levels of reliability and validity. Bowling (1997) suggests that they should range between 0.85 and 0.94. However, there is no agreement over the minimum acceptable standards for scale reliability as some regard 0.5 and others 0.7 as the minimum (Bowling, 1997). Streiner and Norman (1995) suggest 0.85 for internal consistency and 0.5 for stability i.e. test-retest, inter-rater, intra-rater. For validity assessment, low correlations do not necessarily indicate lack of validity (Streiner and Norman, 1995) as the problem could be with the scales or the theory. A correlation of 0.5 would be considered acceptable (Bowling, 1997). Despite these suggestions, care needs to be taken when evaluating instruments that combinations of causal and indicator variables because the psychometric methods lead to poor judgement of validity and reliability (Fayers and Hand, 2002). Fayers and Hand (2002) demonstrate that where scales contain causal variables, correlations can be expected to be low.

3.4 Practicality

This refers to feasibility and acceptability of the instrument. The terms feasibility and acceptability have been used interchangeably by different authors (Mohide et al. 1988: Hurny et al. 1998: Unic et al. 1998: Perez et al. 1997) to refer to similar concepts. Acceptability refers to whether subjects regard the questions and the valuation task acceptable to them. This may concern whether the questions are offensive, ethical, who should be asked the questions (e.g. in the case of terminally ill, how should the questions be framed?). Acceptability can be reflected in, whether the subjects are prepared to contemplate the exchanges used in the valuation questions. For example trading off life years, imagining or contemplating death, or even comparing healthy people with unhealthy

people. Another aspect of acceptability is comprehensibility referring to whether the subjects have a good understanding of the questions. Feasibility refers to whether the method is capable of being carried out in practice and may include issues like method of administration, cost in terms of time. Practicality has been judged through assessment of ease of administration, completion rates and time and comprehension in addition to a host of opinion questions regarding the measures (see table 3.2).

APPENDIX 4.1: LITERATURE SEARCH TERMS

SEARCH TERMS

Broad search area	Search terms
Disease (condition) specific instruments (outcomes or measures)	• All disease categories (see listed below) • Diseas*, disease* • Condition* • Patient* • Instrument* • Outcome* • Measure* • Condition specific measure* • Disease specific measure* • Disease instrument* • Condition specific instrument*
Utilities, preferences and values	• valu*, utilit* preference* • Utilit* measure* • utilit* scal* • Patient* value* • Preference* • Disease* • Condition* • Patient*
Valuation techniques and health state valuation	• Visual analog* scal*, VAS, • Time trade off, time-trade off, TTO, • Standard gamble, SG • Magnitude estimation, ME • Person trade off, person trade-off, PTO • Valuation • Valuation technique* • Health stat* valuation*, health status • 'Quality-adjusted-life-years', 'quality –of-life', 'health-care-and-economics-organizations', 'health-status', 'health-status-indicators', 'outcome-assessment-health-care' • economic evaluat* • cost effective analys* • cost utilit* analys*

DISEASE CATEGORIES (All MeSH Categories)

- Bacterial infections and mycoses
- Cardiovascular diseases
- Congenital, hereditary and neonatal diseases and abnormalities
- Digestive system diseases
- Endocrine diseases
- Eye diseases
- Female genital diseases and pregnancy complications
- Hemic and lymphatic diseases
- Immunologic diseases
- Musculoskeletal diseases
- Neoplasms
- Nervous system diseases
- Nutritional and metabolic disease
- Otorhinolaryngologic diseases
- Parasitic diseases
- Pathological conditions, signs and systems
- Respiratory tract diseases
- Skin and connective tissue diseases
- Stomatognathic diseases
- Urologic and male genital diseases
- Virus diseases

APPENDIX 4.2: REVIEW QUESTIONS

1. Background details

- Year study was undertaken
- Country of study and of the first author
- What diseases or conditions were studied for measurement and valuation?
- Who was the study subjects in measurement and valuation?
- How big were the sample sizes?
- What measurement instruments were used and for what diseases?
- What valuation instruments were used and for what diseases?
- What reasons are given for choice of measurement and valuation instruments?
- Were the instruments used in settings other than those they were developed?
- What were the study objectives?
- What rationale was provided for undertaking the study?

2. Methodological and empirical issues (for measurement and valuation instrument)

- Study aims both methodological or others.
- Which methodological aspects are assessed e.g. validity or reliability etc.?
- What were the validity related aims?
- What types of validity are assessed and how are they assessed?
- What are the empirical findings with regard to different forms of validity?
- What were reliability related aims?
- What types of reliability were assessed and how were they assessed?
- What were the empirical findings with regard to different types of reliability?
- What types of equivalence if any were recognized, assessed and how were they assessed?
- If instruments were adapted, were cross-cultural issues recognized, and if so how were they addressed.
- What methods were followed to do the cross-cultural adaptation?
- What aspects of practicality were evaluated and how were they evaluated?
- What methods and processes were followed to construct the measure for disease outcomes and what types of outcomes are measured?
- What are the different ways validity, reliability and practicality are conceptualized and determined with respect to measurement and valuation instruments?
- What is the content (in terms of domains) of the scenarios valued and how is it arrived at and justified?
- Whose values are used in these studies and what is the justification provided?
- A listing of factors affecting values if assessed in the study
- How the disease specific utilities were used (i.e. intended, implied or explicit use the utilities were put to) and whether this was in EE or HCDM.
-

APPENDIX 4.3: REVIEWED PAPERS

1. Bennett, K.J., Torrance, G.W., Boyle, M.H. and Guscott, R. 2000. Cost-Utility Analysis in Depression: The McSad Utility Measure for Depression Health States. *Psychiatric Services. 2000, 51: 1171-1176.*

2. Sherbourne, C.D., Keeler, E., Unutzer, J., Lenert, L. and Wells, K.B. 1999. Relationship Between Age and Patients' Current Health State Preferences. *Gerontologist. 1999, 39(3): 271-278.*

3. Brown, G.C. 1999. Vision and Quality of Life. *Transactions of American Ophthalmological Society. 1999, 97: 473-511.*

4. Revicki, D.A. and Wood, M. 1998. Patient-Assigned Health State Utilities for Depression-Related Outcomes: Differences by Depression Severity and Antidepressant Medications. *Journal of Affective Disorders. 1998, 48: 25-36.*

5. Brown, G.C., Brown, M.M., Sharma, S., Brown, H., Gozum, M. and Denton, P. 2000. Quality of Life Associated with Diabetes mellitus in an Adult Population. *Journal of Diabetes complications. 2000. 14: 18-24*

6. Jacobson, J.J., Maxson, B.B., Mays. K. and Kowalski, J. 1992. A utility Analysis of Dental Implants. *International Journal of Maxillofacial Implants. 1992. 7: 381-388.*

7. Rutten-van Molken, M.P.M.H., Bakker, C.H., van Doorslaer, E.K.A., and van der Linden, S. Methodological Issues of Patient Utility Measurement: Experience fro Two clinical Trials. *Medical Care. 1995, 33(9): 922-937.*

8. Kirigia, J.M. 1998.Economic Evaluation in Schistosomiasis: Valuation of Health States Preferences- A Research Note. *Health Economics. 1998, 7(6): 551-558.*

9. Lee, J.E., Fos, P.J., Zuniga, M.A., Kastl, P.R. and Sung, J.H. 2001. Assessing Health –Related Quality of Life in Cataract Patients: The Relationship between Utility and Health Related Quality of Life Measurement. *Quality of Life Research. 2001, 9: 1127-1135.*

10. Johnston, K., J. Brown, et al. (1998). "Valuing temporary and chronic health states associated with breast screening." *Social Science and Medicine 47(2): 213-22.*

11. Ashby, J., O'Hanlon, M. and Buxton MJ. (1994). " The Time Trade Off Technique: How do the Valuations of Breast Cancer Patients Compare to Those of Other groups?" *Quality of Life Research 3: 257-265.*

12. Chen, A. Y., Daley, J. and Thibault, G.E. 1996. Angina Patient's Ratings of Current Health and Health without Angina: Associations with Severity of Angina and Comorbidity. *Medical Decision Making. 1996, 16: 169-177.*

13. Bosch, J.L. and Hunink, M.G.M. 1996. The Relationship Between Descriptive and Valuational Quality-of-Life Measures in Patients with Intermittent Claudication. *Medical Decision Making. 1996, 16: 217-225.*

14. Swan, J.S., Fryback, D.G., Lawrence, W.F., Sainfort, F., Hagenauer, M.E. and Heisey, D.M. 2000. A Time-Tradeoff Method for Cost-Effectiveness Models Applied to Radiology. *Medical Decision Making. 2000, 20: 79-88.*

15. Swan, J.S., Fryback, D.G., Lawrence, W.F., Katz, D.A., Heisey, D.M., Hagenauer, M.E., Ellzer, P.M. and Jacobson, B.K. 1997. MR and Conventional Angiography: Work in Progress Toward Assessing Utility in Radiology. *Academic Radiology. 1997, 4: 475-482.*

16. Havranek, E.P., Mcgovern, K.M., Weinberger, J., Brocato, A., Lowes, B.D. and Abraham, W.T. 1999. Patient Preferences fore Heart Failure Treatment: Utilities are Valid Measures of Health Related Quality of Life in Heart Failure. *Journal of Cardiac Failure. 1999. 5(2): 85- 91.*

17. Bakker, C., Rutten, M., van Santen-Hoeufft, M., Bolwijn, P. van Doorslaer, E., Bennett, K. and van der Linden, S. 1995. Patient Utilities in Fibromyalgia and the Association with Oter Outcome Measures. *Journal of Rheumatology. 1995. 22:1536-43.*

18. Guyatt, G.H., King D.R, Feeny, D. H., Stubbing, D., and Goldstein, R.S. 1999. Generic and Specific Measurement of Health Related Quality of Life in a Clinical Trial of Respiratory Rehabilitation. *Journal of Clinical Epidemiology, 1999. 52(3): 187-192.*

19. Gabriel, S.E., campion, M.E. and O'Fallon, W.M. 1994. A Cost-Utility Analysis of Misoprostol Prophylaxis for Rheumatoid Arthritis Patients Receiving Nonsteroidal Antiinflamatory Drugs. *Arthritis and Rheumatism. 1994. 37(3): 333-341.*

20. Yee, G.C. 1997. Cost-Utility Analysis of Taxane Therapy. *American Journal of Health System Pharmacy. 1997, 54(suppl. 2):S11-15.*

21. Mackeigan, L.D., O'Brien, B. and Oh, P.I. 1999. Holistic Versus Composite Preferences for Lifetime Treatment Sequences for Type 2 Diabetes. *Medical Decision Making. 1999, 19: 113-121.*

22. Chen, S., Shaheen, A. and Garber,, A. 1998. Cost Effectiveness and Cost-Benefit Analysis of Using Methotrexate Vs Geockerman Therapy for Psoriasis: A Pilot Study. *Archives of Dermatology. 1998. 134: 1602-1808.*

23. Kerrigan, C.L., Collins, E.D., Kneeland, T.S., Voigtlaender, D., Moncur, M.M., Matheney, T.H., Grove, M.R. and Tosteson, A.N.A. 2000. Measuring Health State Preferences in Women with Breast Hypertrophy. *Plastic Reconstructive Surgery. 106: 280-8.*

24. Douzdjian, V., Ferrara, D. and Silvestri, G. 1998. Treatment Strategies for Insulin-Dependent Diabetics with ESRD: Cost-Effectiveness Decision Analysis Model. *American Journal of Kidney Diseases. 31(5): 794-802.*

25. Baltussen, R.M.P.M., Sanon, M., Sommerfeld, J. and Wurthwein, R. 2002. Obtaining Disability Weights in Rural Burkina Faso using a Culturally Adapted Visual Analog Scale. *Health Economics, 2000. 11(2): 155-164.*

26. Gerard, K., Johnston, K. and Brown, J. 1999. The Role of a Pre-Scored Multi-Attribute Health Classification Measure in Validating Condition-Specific Health State Descriptions. *Health Economics. 1999, 8: 685-699.*

27. Gabriel, S.E., Kneeland, T.S., Melton, J.L., Moncur, B.E. and Tosteson, A.N.A. 1999. Health-Related Quality of Life in Economic Evaluations for Osteoporosis: Whose Values Should we Use? *Medical Decision Making. 1999, 19: 141-148.*

28. Bombardier, C., Ware, J., Russell, I.J., Larson, M., Chalmers, A. and Read, J..L. 1986. Auranofin Therapy and Quality of Life in Patients with Rheumatoid Arthritis: Results of a Multi-Center Trial. *The American Journal of Medicine. 1986, 81(4): 565-578.*

29. Kirsch, J. and McGuire, A. 2000. Establishing Health State Valuations for Disease Specific States: An Example from Heart Disease. Health Economics, 2000. 9: 149-158.

30. Sauraez-Almazor, M.E. and Corner-Spady, B. 2001. Rating of Arthritis Health States by Patients, Physicians, and the General Public. Implications for Cost Utility Analysis. *Journal of Rheumatology, 2001. 28; 648-56.*

31. Hayman, AJ.A., Fairclough, D.L., Harris, J.R., and Weeks, J.C. 1997. Patient Prefernces Concerning the Trade-Off between te Risks and Benefits of Routine radiation Therapy after Conservative Surgery for Early Stage Breast Cancer. *Journal of Clinical Oncology, 1997. 15:1252-1260.*

32. Stolk, E.A., Busschbach, J.J.V., Caffa, M., Meuleman, E.J.H., Rutten, F.F.H. 2000. Cost Utility Analysis of Sildenafil Compared with Papaverine-Phentolamine injections. *British Medical Journal. 2000. 320: 1-6.*

33. Blumenschein, K. and Johannesson, M. 1998. Relationship Between Quality of Life Instruments, Health State utilities, and Willingness to Pay in Patients With Asthma. *Annals of Allergy, Asthma and Immunology, 1998. 80: 189-94.*

34. Ackerman, S.J., Rein, A.L., Blute, M., Beusterien, K., Sullivan, E.N., Tanio, C.P., Manyak, M.J. and Strauss, M.J. 2000. Cost Effectiveness of Microwave Thermotherapy in Patients with Benign Prostatic hyperlasia: Part 1-Methods. *Urology, 2000. 56: 972-980.*

35. Zug, K.A., Littenberg, B., Baughman, R.D., Kneeland, T., Nease, R.F., Sumner, W., O'Connor, G.T., Jones, R., Morrison, E. and Cimis, R. 1995. Assessing the Preferences of Patients with Psoriasis: A Quantitative, Utility Approach. *Archives of Dermatology, 1995. 131; 561-568.*

36. Souchek, J., Stacks, J.R., Brody, B., Ashton, C.M., Giesler, B., Byrne, M.M., Cook, K., Geraci, J.M. and Wray, N.P. 2000. A Trial for Comparing Methods for Eliciting Treatment Preferences from Men with Advanced Prostate Cancer: Results from the Initial Visit. *Medical Care, 2000. 38(10): 1040-1050.*

37. Bayoumi, A M. and Redelmeier, D.A. 1999. Economic Methods for Measuring the Quality of Life Associated with HIV Infection. *Quality of Life Research, 1999. 8: 471-480.*

38. McLachlan, S., Pintilie, M. and Tannock, I.F. 1999. Third line Chemotherapy in Patients with Metastatic Breast Cancer: An Evaluation of Quality of Life and Cost. *Breast Cancer Research and Treatment, 1999. 54:213-223.*

39. Miller, A.R., Cantor, S.B., Peoples, G.E., Pearlstone, D.B. and Skibber. J.M. 2000. Quality of Life and Cost Effectiveness Analysis of Therapy for Locally Recurrent Rectal Cancer. *Diseases of Colon and Rectum, 2000. 43: 1695-1703.*

40. Matchar, D.B., McCrory,D.C. and Bennett, C.L. 1997. Treatment Considerations for Persons with Metastatic Prostate Cancer: Survival Versus Out-of-Pocket Costs. *Urology, 1997. 49: 218-224.*

41. Dranitsaris, G., Leung, P., Mather, J. and Oza, A. 2000. Cost-Utility Analysis of Second Line Hormonal Therapy in Advanced Breast Cancer: A Comparison of Two Aromatase Inhibitors to Megestrol Acetate. *Anti-Cancer Drugs, 2000. 11:591-601.*

42. Leung, P.P., Tannock, I.F., Oza, A.M., Poudziunas, A. and Dranitsaris, G. 1990. Cost-Utility Analysis of Chemotherapy using Paclitaxel, Docetaxel, or Vinorelbine for Patients with Anthracycline-Resistant Breast Cancer. *Journal of Clinical Oncology. 1999, 17(10): 3082-3090.*

APPENDIX 4.4: ASSESSMENT OF CONSTRUCT VALIDITY: VALUATION INSTRUMENTS

Author / Year/ [country/ condition/ sample size/ valuation instruments]A	Validity related aims	Hypothesized relationships	Analytical methods for testing construct validity	Empirical findings on construct validity
Jacobson et al, 1992 [Canada; dental implants; n=111patients; VAS]	None mentioned	Testing the hypothesis that implant patients would score lower on the rating scale immediately post-surgery before prosthesis insertion than at six months after prosthesis insertion.	Comparing Feeling Thermometer rating scores 1 month before prosthesis and 6 months after using one tailed *t* test	Pre-insertion mean score was lower than post-insertion score. P value was less than 0.05. Hence RS was valid.
Havranek et al, 1999 [USA; Heart Failure; n=50; TTO and VAS]	To evaluate construct validity for utilities.	They hypothesized that utilities would have significant relationships with 6-minute distance walk, standard HRQL questionnaire results and a visual analogue score of overall health (direction not initially stated).	Assessed by testing for expected relationships between utilities and other measures of HRQL, overall health status, and physical performance using correlation and $P<0.01$	There were significant relationships between utilities and other measures of HRQL, overall health status, and physical performance. Hence utilities have validity as measures of HRQ, such that higher scores correspond to less severe disease.

Bakker et al, 1995 [Nethrlands; Arthritis; n=73; SG and VAS]	To evaluate construct validity of RS and SG methods	Improvements in utilities were expected to be associated with improvements in scores for global health, SIP, AIMS, mHAQ, pain and stiffness.
Guyatt et al, 1999 [Canada; Chronic airflow limitation; n=89; SG]	To compare the validity and responsiveness of HRQL measures	Correlation between generic and disease specific measures as well as between generic measures and assessed based on a priori expectations of the authors.

• Construct validity of MUMQ was tested by Spearman correlation coefficient between baseline utility values and baseline scores for global health, SIP, AIMS, mHAQ, pain and stiffness. • Discriminant validity or sensitivity to change was tested in two ways. One using Spearman correlation coefficient between changes in utility values and changes in other outcome measures. Two, multiple regression analysis (stepwise forward) with changes in utility scores as dependent variables and treatment and changes in other health outcomes as independent variables. • Sensitivity to change assessed by efficiency as mean change of the measure divided by the standard deviation of change $P< 0.05$ (one tail test)	RS: Construct validity of RS utilities was supported by significant correlations with measures of global health, pain, SIP, AIMS, mHAQ. Changes in rating scale scores correlated significantly with changes in four dimensions of AIMS, SIP, pain and global health while in SG significant correlations were with Some dimensions of AIMS, and global health. RS: Has better discriminant validity than SG, hence more sensitive to change. Changes in RS scores could be explained to a higher degree than changes in SG SG: lower construct validity as it correlated considerably less with global health, SIP, AIMS, mHAQ, pain and stiffness.
Pearson's correlation s $P<0.05$	Correlation between SG, QWB and disease-specific measures were uniformly weak (0.2-0.35) and very weak (<0.2) with SIP.

Gabriel et al, 1994 [USA; rheumatoid arthritis; n=57; VAS and TTO]	None stated	Construct validity based on expected ranking a priori of the health state scenarios. (The authors keep referring to these expectations but have not set them out explicitly).	• Comparisons of utility values for different states • Convergent validity assessed by consistency of results between VAS and TTO (not explained how this was done. The study does not set out the methods used to test validity but presents the results suggesting that these methods were used. They only say that 'we documented the face, content, construct and convergent validity' without outlining clearly how).	Demonstrated by worse of scenarios getting lower ranking and utility scores. Utility score for 'no ulcer' was significantly higher than 'ulcer requiring medical treatment' which was significantly higher than that for 'ulcer requiring surgery'. 'Prophylaxis without side effects' was scored higher than 'prophylaxis with side effects'. Convergent validity suggested by the consistency of results with the two measurement techniques.
Kerrigan et al, 2000 [USA; Breast hypertrophy, n=47; SG and VAS]	To validate a utility measure for estimating health related quality of life in women with breast hypertrophy.	The worse off the perceived health state of the patient the lower the utility value	Evaluated by comparing utility scores in women with hypertrophy to three sub-scales of SF-36 health survey namely physical function, role physical function and body pain (they have been shown to be significantly affected by breast hypertrophy). Comparisons use Wilcoxon rank-sum tests.	Patients who reported greater bodily pain on SF-36 also reported lower utilities than did those who reported less pain. Women with poor physical function and role physical function had lower utility scores. Women with breast hypertrophy had lower utilities for current health than those without for Table but not Wheel. For both Table and Wheel, those rating their health (SF-36) as excellent or very good had higher utilities than those rating it as good fair or poor.

Baltussen et al, 2002 *[Burkina Faso; 9 diseases; n=56; VAS]*	*To evaluate the validity of a culturally adapted VAS.*	*None stated.*	Construct (convergent) validity studied by a comparison with the results of implicit rank order *following from final evaluation with an explicit* rank order exercise carried out after final evaluation using Spearman rank correlation coefficient (only referred to in the results section and not discussed in previous sections. Therefore not easy to interpret what this means)	The Spearman rank correlation coefficient was 0.86 and 0.94 for *lay people and health professionals.* Therefore both panels were consistent in their evaluations indicating construct validity.
Gerard et al, 1999 [UK; Breast cancer; n=440; TTO]	To learn more about the convergent validity of direct and indirect approaches to valuation.	None stated	Convergent validity: Assessed by using a measure of the extent of agreement between values using Bland and Altman's method of measuring agreement. Paired ranks were also used to test for convergent validity of ranks between TTO and EuroQoL values.	Agreement between mean EuroQoL value and TTO value was low. Convergent validity of the two methods was poor. There was considerable variation to values attached to different descriptions using direct (TTO) and indirect approach (EuroQoL).
Souchek et al, 2000 [USA; prostate cancer; n=120; RS, TTO and SG]	To test the convergent validity of the SG, TTO and RS techniques for estimating individual and group values for specific health states.	None explicitly stated	• Comparison of utility values with the rank orders given to health states and convergence among techniques. • Pearson's correlation coefficients for utility values between techniques were calculated for each participant to measure the amount of convergence between techniques. • Correlation between rank order of health states and ordering by utility values using inconsistency and differentiation index to assess validity	Pearson correlation coefficients between techniques was moderate to high (RS/SG-0.74; TTO/SG-0.69; RS/TTO- 0.76) indicates convergent validity for the 3 methods for eliciting utility scores. There was high level of inconsistency in individual scores and poor differentiation of individual states for all methods.

APPENDIX 5.1: Intestinal Parasitic Infections related impacts, *indicators and methods of assessment*

study	(n), study area and age group	Impact related aim	Study design	Indicators	Findings
(Lawless 1994)	(87) Kenya (coast) 6-11 years	Effects of iron supplementation on appetite and growth	Randomized double blind placebo controlled trial	• Anthropometry: (nutritional status) weight, height, triceps, and subscapular skin fold. • Appetite: ad libitum consumption of a mid morning snack, subjective assessment of own appetite. • Physical examination for morbidity screening: spleen size, liver enlargement, fecal exam for intensity and prevalence of infection (epg), hemoglobin and serum ferratin to screen for iron nutritional status.	• Iron supplements resulted in improved growth and appetite. • There was high prevalence of anaemia and high levels of IWI. • There was significant increase in weight, height and other anthropometric measures for treated children.
(Stoltzfus 1996)	(203) Zanzibar Grade 1-4	Relationship between hookworm infection, blood loss and iron status	Systematic selection of hookworm infected children	• Fecal egg counts • Haemoglobin and serum ferratin concentration • Fecal heme (for intestinal blood loss)	• Intestinal blood loss strongly and linearly related to hookworm egg counts. • Higher levels of blood loss associated with worse iron status.
(Tanumihardjo 1996)	(309) Indonesia 0.6-6.6 years	Effect of vitamin A supplementation and treatment for ascaris infection on vitamin A status.	Randomized control trial	• Serum retinol concentrations • Modified relative dose response to assess vitamin A status. • Fecal samples analysis • Physical examination • Anthropometric evaluation.	• Vitamin A supplementation was most important in improving status. • Ascaris infection treatment and treatment combined with vitamin A supplement were not significant. • Deworming had neither beneficial nor detrimental effect on vitamin A utilization.

(Raj 1996)	(246) Malaysia School children	Influence of IWI on intestinal permeability (mucosal damage, altered small intestinal mucosal injury.	2 schools in 2 settings, compared infected with non-infected	• Anthropometric: height, weight • Fecal sample analysis	• Intestinal helminths in particular ascaris was associated with altered intestinal permeability. • Intestinal worms infection exerts minor influence on intestinal permeability.
(Watkins 1996)	(246) Guatemala 7-12 years	Effects of deworming on indicators of school performance, physical growth and cognitive function.	Randomized double blind study (stratified by gender and age).	• Tests of reading and vocabulary: (PPVT) Pea body picture vocabulary test (interamerica vocabulary and reading tests and Spanish version of peabody picture vocabulary test), school attendance. • Fecal samples analysis.	• No positive effect of deworming • A higher worm burden was associated with poorer performance. • Effects of worms are likely to be modest, compared to social factors like gender and culture.
(Kightlinger 1996)	(663) Madagascar 4-10 years	Relationship between Ascaris, worm burden, growth status, general delayed cuteneous hypersensitivity response and plant anthelmintic use.	Prospective 12 months survey	• Fecal sample analysis • Anthropometric: weight, height, triceps skinfold, upper arm circumference (to assess growth status and growth velocity). • Clinical examinations • Worm expulsion and counts • Skin tests	• Worm intensity and growth status were inversely correlated. • Number of worm inversely correlated with poor growth indices
(Hadju, Stephenson et al. 1996)	(75) Indonesia 6-10 years	Appetite and growth in school boys with IWI	RCT (follow-up 3 and 7 weeks later)	• Anthropometry: weight, age, • Appetite: ad libitum consumption of a midmorning snack combined with subjective assessment of own appetite. • Fecal sample analysis for prevalence and intensity of infection.	• Treated group exhibited significantly greater in weight and percentage weight for age. • Appetite increased • High egg and prevalence of ascaris reduction rates. • Faster growth in the treated group • Treatment of ascaris infection was associated with increases in appetite and weight gain in children. • Treatment for hookworm infection in combination with

					ascaris and trichuris and schistosomiasis improved appetite and growth in Kenya school children.
(Stoltzfus 1997a)	(3063) Zanzibar Grade 1-5 Average age 10 years	Effect of deworming on growth	Randomized control trial	• Growth status: weight, height, age (height for age, weight for age and weight for height, body mass index) • Fitness • Iron status: hemoglobin, erythrocyte protoporphyrin, serrum ferratin • Fecal and urine samples analysis	• Deworming improved growth of school children though the improvements were small. • Growth, appetite, fitness and iron status improved with treatment with anthelmitics. • Helminthes control is necessary but not sufficient to substantially improve the growth of school children in areas such as Zanzibar.
(Stoltzfus 1997b)	(3595) Zanzibar 7-13 years	To describe the distribution of iron deficiency and anemia and their determinants in school children.	Multivariate analysis	• Hemoglobin, crythrocyte protoporphyrin and serum ferratin concentrations • Fecal samples analysis • Nutritional indicators, height, eight, age	• Anemia was strongly linked to hookworm status or intensity • Anthelmintic therapy is essential component of anemia control where hookworm is endemic.
(Stephenson 1990)	(36) Kenya 6-12 years	To determine physical fitness before and after treatment for helminthes infestations	Randomized Controlled trial (7 wks follow-up)	• Harvard step test: fitness score, heart rate • Fecal samples analysis • Anthropometric measures: weight, height, arm circumference, triceps and sub-scapular skin fold thickness. • Hemoglobin •	• Anthelmintic treatment improves physical fitness • Treatment of soil transmitted helminths in children may improve the ability to perform work • Anemia and low weight for height (wasting) in male workers inversely related to physical fitness and productivity (citation from a previous study).
(Stephenson 1994)	(53) Kenya 7-13 years	To determine effect of treatment on physical fitness, parasite rates, growth rates, appetite and	Randomized Controlled trial (4 months follow-up)	• Harvard step test: fitness score, heart rates and resting heart rates (indicator of physical fitness). • Fecal sample analysis • Anthropometric measures: weight, height, arm circumference, triceps and	• Anthelmintic treatment improved physical fitness, growth and appetite. • Treatment of helminths infection resulted in significant improvements in growth,

		hemoglobin concentration on children with ascariasis, trichuiriasis and hookworm infection.		sub-scapular skin fold thickness. • Appetite: Ad libitum consumption of mid morning snack and subjective own appetite assessment.	physical fitness, resting heart rates, perceived appetite and small but significant improvements in measured appetite and hemoglobin concentrations. • Children and adults grow and feel better, are healthier, more physically and mentally active and productive after treatments for helminthes infestations and anemia.
(Adams 1994)	(55) Kenya 5-10 years	Impact of treatment with albendazole on growth, physical activity level and appetite and their interrelationships	Randomized controlled trial (9 weeks)	• Fecal samples analysis • Physical activity level (measured during free play with a motion recorder) • Appetite: Children's own rating • Anthropometric measures: weight, height, arm circumference, triceps and sub-scapular skin fold thickness. • Hemoglobin	• Treatment of undernourished children for intestinal helminthes infestations with albendazole may improve growth and appetite and increase spontaneous physical activity. • Treatment of helminths leads to increased free play activity by children (demonstrates improvements in quality of children's lives, potential for increased productivity and increased ability to participate in learning activities.
Parker, M., 1992	(22) Sudan Women infected with schistosomiasis	Whether schistosome infection is sufficiently debilitating to alter daily activity patterns among women engaged in agricultural work in the cotton-picking season.	Paired case control study	• Minute by minute recording of daily activities: picking cotton, domestic chores and personal care.	• Infected women were less likely to participate in those activities which did not directly affect or contribute to their domestic and agricultural productive output • Infection with schistosoma mansoni was debilitating and interfered with daily activities.

APPENDIX 5.2: Definitions of Intensity of S. Mansoni infection used in selected studies (epgf)

Study/country	Light	moderate	heavy
Stephenson, 1987	1-100	101-400	>400
	1-99	100-499	>500
Mohamed-Ali et al. 1991	1-100	100-400	> 401
Jaoko et al. 1996	1-100	101-800	>800
WHO, 1985			101-800
WHO, 1993	<100	101-400	>400

Source: cited studies. NB//: see Stephenson, 1987. P. 227 for values from different studies and for other worms.

APPENDIX 5.3 SCHISTOSOMIASIS MANSONI HRQL QUESTIONNAIRE (English version)

SCHISTOSOMIASIS HRQL QUESTIONNAIRE

(Questionnaire for patients)

[This questionnaire must be completed in pencil]

1	Questionnaire number		

		Day		Month		Year			
2	Date of interview.								

3. Study number and phase ***[Tick the appropriate box]***

1	2	3	4
S1P1 (Pre-test)	S1P2A (Patient study)	S1P2B (Community study)	S2 (Valuation study)

4.	Patient number		

4a. Name of patient_____

Recruited at	1	Kimbimbi Health Center
	2	Nguka Dispensary

5. Name of interviewer_____

6.

Location *[Circle one only]*	1. Tebere
	2. Mutithi
	3. Thiba
	4. Nyangiti
	5. Murinduko

7. Sub-location *[Circle one only]*

1	Kiarukungu	12	Wamumu
2	Mahigaini	13	Thiba
3	Gathigiriri	14	Mathangauta
4	Rukanga	15	Kirimara
5	Kinyaga	16	Ndomba
6	Mathigaini	17	Nyangati
7	Kombuini	18	Riagicheru
8	Kathiga	19	Mugabaciura
9	Kiandegwa	20	Miuu
10	Kabiriri	21	Kamunyange
11	Nguka		

INTRODUCTION

I am (*say your name)* working with Mercy Mugo, a student researcher from the University of Nairobi. We would like to learn from you, your views about of how bilharzia (schistosomiasis) affects health and quality of life of people. You have been selected to participate in the study on account of your experience with the illness. Your views and opinions are important to a clear understanding of the impact the illness has on people's quality of life.

I am here today to ask for your help by requesting you to give me an interview. Can we go on? Thank you.

(If yes, present the information sheet and consent form and allow the respondent to read through. If unable to read, read out for her/him. Ask the respondent to sign the consent form, before continuing with the interview).

[Interview starts]
Record start time ..... Morning..... Afternoon .....{***Tick as appropriate***]

SECTION A: HISTORY OF ILLNESS
In this section I will ask you questions related to your recent history of illness. In answering these questions think of the last 2 weeks
[Circle the correct response.]

8. In the last two weeks, did you experience any illness?
   1. Yes ***[go to9].*** 2. No ***[go to11].*** 3. DN ***[go to11]*** 4. NR ***[go to11]***

9. Did you know what illness you suffered from?
   1. Yes ***[go to 10].*** 2. No ***[go to 11].*** 3. DN ***[go to11*** 4. NR ***[go to11]***

10. What was the illness? _____

11. In the last two weeks, did you experience any health problems?
    1. Yes ***[go to 12].*** 2. No ***[go to 17].*** 3. DN ***[go to17]*** 4. NR ***[go to17]***
12. Did you know what health problems you suffered from?
    1. Yes ***[go to 13].*** 2. No ***[go to14].*** 3. DN ***[go to14]*** 4. NR ***[go to14]***
13. What were the health problems? _____

>> **Prompt:** *Was there anything else?**[record all the illnesses and health problems the respondent mentions, each on a separate line. Use additional space below if needed]***

1._____
2._____
3._____

14. As a result of the illness, were you restricted in carrying out any of your daily duties?
    1. Yes ***[Ask 15 and 16].*** 2. No ***[go to 17].*** 3. DN ***[go to17]*** 4. NR ***[go to17]***

15. Please tell me all the kinds of daily duties you were restricted in.***[list each activity on a separate row. Use additional space if required]***

	Activity restricted
15a	
15b	
15c	
15d	
15e	

16. Please tell me the number of days you were restricted in carrying out your daily duties. ***[Go back to responses in Q15 above and fill in the extent of restriction for each]***

Activity restricted	**< 1 day**	**1-3 days**	**4-6 days**	**1-2 weeks**
16a				
16b				
16c				
16d				
16e				

17. Are you having any illness now?
    1. Yes ***[go to 18].*** 2. No ***[go to 20].*** 3. DN ***[go to20]*** 4. NR ***[go to20]***
18. Do you know what illness you are suffering from now?
    1. Yes ***[go to 19].*** 2. No ***[go to 20].*** 3. DN ***[go to 20]*** 4. NR ***[go to20]***
19. Which illness?_____
20. Are you having any health problems now?
    1. Yes ***[go to 21].*** 2. No ***[go to 23].*** 3. DN ***[go to23]*** 4. NR ***[go to23]***
21. Do you know what health problems you are suffering from now?
    1. Yes ***[go to 22].*** 2. No ***[go to 23].*** 3. DN ***[go to23]*** 4. NR ***[go to23]***

22. What health problems? _____
23. Have you ever suffered from bilharzia (schistosomiasis mansoni)?
    1. Yes *[go to 24].* 2. No *[go to 34].* 3. DN *[go to34]* 4. NR *[go to34]*
24. When was your last illness episode with bilharzia (schistosomiasis)?
    Month__ __ Year __ __ __ __ *[Go to 25]*
25. Did you seek medical assistance?
    1. Yes *[go to 26].* 2. No *[go to 33].* 3. DN *[go to33]* 4. NR *[go to33]*

26	Did you seek assistance at a GOK facility?	Yes *[go to 27]*	No *[go to28]*	DN *[go to 28]*	NR *[go to28]*

27. Which GOK facilities did you go to? ***[Circle all the facilities visited and indicate the name of the facility]*** ***>>PROMPT:*** *Ask the respondent to tell you the order in which they went to different facilities and rank as 1,2 3, ... from 1st to the last.*

code	Facility	Ranking of facility by order of visit
1	Kenyatta National Hospital	
2	Provincial Hospital	
3	District hospital	
4	Health Center	
5	Dispensary	

28	Did you seek assistance at a Non-GOK facility?	Yes *[go to29]*	No *[go to 30]*	DN *[go to 30]*	NR *[go to30]*

29. Which Non-GOK facilities did you go to? ***[Circle all the facilities visited and indicate the name of the facility]*** *>> Ask the respondent to tell you the order in which they went to different facilities and rank as 1,2 3, ... from 1st to the last.*

code	Facility	Ranking of facility by order of visit
1	Private hospital	
2	Mission Hospital	
3	Mission dispensary	
4	Nursing home	
5	Private clinic	

30	Did you seek assistance from a traditional healer?	Yes *[go to31]*	No *[go to 33]*	DN *[go to 33]*	NR *[go to 33]*
31	Did you seek assistance from a herbalist?	Yes *[go to32]*	No *[go to 33]*	DN *[go to 33]*	NR *[go to 33]*
32	Did you seek assistance from other sources? (specify)	Yes *[go to34]*	No *[go to 33]*	DN *[go to 33]*	NR *[go to 33]*

33. How did you cope with the illness? _____

Please describe in your own words how you generally feel about your health state at present? _____

>>***PROMPT:*** *Ask the respondent what they are thinking of when they talk of health state* ***[record the responses below]***

_____

If we assign being dead 0 and life without any health problems 100, where, on that scale would you place your own health state at present? Dead

Life without any health problem

PROP:

SECTION B: SYMPTOMS

In the next set of questions I would like to ask you about symptoms you may have experienced in the last two weeks.

I will now go through a list of symptoms. Please tell me which of these symptoms you had in the last two weeks.

[Tick one box only against yes, no, NA (not applicable) and NR (none response)]

		YES	NO	NA	NR	
36	In the last two weeks did you have any abdominal pain and discomfort?	*[Go to 37]*	*[Go to 37]*	*[Go to 37]*	*[Go to 37]*	*Ask 47-51*
37	>> PROMPT: I am now going to ask you about different types of diarrhea, like plain diarrhea, watery diarrhea and bloody mucoid diarrhea. Please tell me the type of diarrheas you had.<< Tick the diarrhea that fits the description as appropriate under Q37-39. In the last two weeks did you have any diarrhea?	*[Go to 38]*	*[Go to 38]*	*[Go to 38]*	*[Go to 38]*	*Ask 52-56*
38	In the last two weeks did you have any Watery diarrhea?	*[Go to 39]*	*[Go to 39]*	*[Go to 39]*	*[Go to 39]*	*Ask 57-61*
39	In the last two weeks did you have any bloody mucoid diarrhea?	*[Go to 40]*	*[Go to 40]*	*[Go to 40]*	*[Go to 40]*	*Ask 62-66*
40	In the last two weeks did you have any tiredness?	*[Go to 41]*	*[Go to 41]*	*[Go to 41]*	*[Go to 41]*	*Ask 67-71*
41	In the last two weeks did you have any nausea?	*[Go to 42]*	*[Go to 42]*	*[Go to 42]*	*[Go to 42]*	*Ask 72-76*
42	In the last two weeks did you vomit blood?	*[Go to 43]*	*[Go to 43]*	*[Go to 43]*	*[Go to 43]*	*Ask 77-81*
43	In the last two weeks did you lose your appetite?	*[Go to 44]*	*[Go to 44]*	*[Go to 44]*	*[Go to 44]*	*Ask 82-86*
44	In the last two weeks did you have any itching skin rash/	*[Go to 45]*	*[Go to 45]*	*[Go to 45]*	*[Go to 45]*	*Ask 87-91*
45	In the last two weeks did you have any fever?	*[Go to 46]*	*[Go to 46]*	*[Go to 46]*	*[Go to 46]*	*Ask 92-96*
46	In the last two weeks did you have any dizziness?	*[Go to 102]*	*[Go to 104]*	*[Go to 104]*	*[Go to 104]*	*Ask97-102*

[Go through the column marked "yes" in questions 36-47 above and for each "yes" answer, tick "X" on the last column. For each box ticked "X" ask the questions indicated]

ABDOMINAL PAIN AND DISCOMFORT

47. You said that in the last two weeks you had abdominal pains and discomfort. Please think of those two weeks and tell me the number of days you had abdominal pain and discomfort. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

48 For the days you had abdominal pain and discomfort, how would you describe the intensity of that pain **at its worst**? ***[Tick one box only]***

1 = mild	2 = moderate	3 = severe	4 = Very severe

49. Did the abdominal pain and discomfort disrupt your daily duties in any way?

1. Yes ***[go to 50].*** 2. No ***[see 37 last col.].*** 3. DN ***[see 37 last col.]*** 4. NR ***[see 37 last col.]***

50. How did abdominal pain and discomfort disrupt your duties daily?

_____

_____

_____

51. For the days you had abdominal pain and discomfort, for how long were your normal activities disrupted? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

DIARRHOEA

52. Please tell me the number of days you had diarrhea in the last two weeks. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

53 For the days you had diarrhea, **at its worst**, how many times per day did you pass stool? ***[Tick one box only]***

1 = 3-9	2 = 9-15	3 = 15-21	4 = over 21

54. Did diarrhea disrupt your daily duties in any way?

1. Yes ***[go to 55].*** 2. No ***[see 38 last col.].*** 3. DN ***[see 38 last col.]*** 4. NR ***[see 38 last col.]***

55. How did diarrhea disrupt your daily duties?

_____

_____

_____

56. For the days you had diarrhea, for how long were your daily duties disrupted? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

WATERY DIARRHOEA

57. Please tell me the number of days you had watery diarrhea in the last two weeks. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

58. For the days you had watery diarrhea, **at its worst**, how many times per day did you pass stool? ***[Tick one box only]***

1 = 3-9	2 = 9-15	3 = 15-21	4 = over 21

59. Did watery diarrhea disrupt your daily duties in any way?

1. Yes ***[go to 60].*** 2. No ***[see 39 last col.].*** 3. DN ***[see 39 last col.]*** 4. NR ***[see 39 last col.]***

60. How did watery diarrhea disrupt your daily duties?

_____

_____

_____

61. For the days you had watery diarrhea, for how long were your daily duties disrupted? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

BLOODY MUCOID DIARRHOEA

62. Please tell me the number of days you had bloody mucoid diarrhea in the last two weeks. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

63 For the days you had bloody mucoid diarrhea, **at its worst**, how many times per day did you pass stool? ***[Tick one box only]***

1 = 3-9	2 = 9-15	3 = 15-21	4 = over 21

64. Did bloody mucoid diarrhea disrupt your daily duties in any way?

1. Yes ***[go to 65].*** 2. No ***[see 40 last col.].*** 3. DN ***[see 40 last col.]*** 4. NR ***[see 40 last col.]***

65. How did bloody mucoid diarrhea disrupt your daily duties?

_____

_____

_____

66. For the days you had bloody mucoid diarrhea, for how long were your daily duties disrupted? ***[Tick one box]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

TIREDNESS

67. Please tell me the number of days you had tiredness in the last two weeks. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

68. For the days you had tiredness, **at its worst**, how tired did you feel? ***[Tick one box only]***

1 = A little tired	2 = Somewhat tired	3 = Very tired	4 = Extremely tired

69. Did tiredness disrupt your daily duties in any way?

1. Yes ***[go to 70].*** 2. No ***[see 41 last col.].*** 3. DN ***[see 41 last col.]*** 4. NR ***[see 41 last col.]***

70. How did tiredness disrupt your daily duties?

_____

_____

_____

71. For the days you had tiredness, for how long were your daily duties disrupted? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

VOMITING BLOOD

77. Please tell me the number of days you vomited blood in the last two weeks. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

78. For the days you vomited blood, **at its worst,** how would you describe the amount of the blood in the vomit? ***[Tick one box only]***

1 = very little	2 = A little	3 = moderate	4 = A lot

79. Did vomiting blood disrupt your daily duties in any way?

1. Yes ***[go to 80].*** 2. No ***[see 43 last col.].*** 3. DN ***[see 43 last col.]*** 4. NR ***[see 43last col.]***

80. How did vomiting blood disrupt your daily duties?

_____

_____

_____

81. For the days you vomited blood, for how long were your daily duties disrupted? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

LOSS OF APPETITE

82. Please tell me the number of days you lost your appetite in the last two weeks. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

83. For the days you lost your appetite, **at its worst**, how much of the food that you normally eat were you able to eat at meal time? ***[Tick one box only]***

1 = About three quarters	2 = Half to a quarter the amount	3 = not more than two spoonfuls	4= None at all

84. Did losing your appetite disrupt your daily duties in any way?

1. Yes ***[go to 85].*** 2. No ***[see 44 last col.].*** 3. DN ***[see 44 last col.]*** 4. NR ***[see 44 last col.]***

85. How did losing your appetite disrupt your daily duties?

_____

_____

_____

86. For the days you lost your appetite, for how long were your daily duties disrupted? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

ITCHING SKIN RASH

87. Please tell me the number of days you had itching rashes on your skin in the last two weeks. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

88. For the days you had itching rashes on your skin, **at its worst**, how would you describe its intensity? ***[Tick one box only]***

1 = mild	2 = moderate	3 = severe	4 = very severe

89. Did itching rashes on your skin disrupt your daily duties in any way?

1. Yes ***[go to 90].*** 2. No ***[see 45 last col.].*** 3. DN ***[see 45 last col.]*** 4. NR ***[see 45 last col.]***

90. How did itching rashes on your skin disrupt your daily duties?

_____

_____

_____

91. For the days you had itching rashes on your skin, for how long were your daily duties disrupted? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

FEVER

92. Please tell me the number of days you had fever in the last two weeks. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

93. For the days you had fever, **at its worst**, how would you describe its intensity? ***[Tick one box only]***

1 = mild	2 = moderate	3 = severe	4 = very severe

94. Did fever disrupt your daily duties in any way?

1. Yes ***[go to 95].*** 2. No ***[see 46 last col.].*** 3. DN ***[see 46 last col.]*** 4. NR ***[see 46 last col.]***

95. How did fever disrupt your daily duties?

_____

_____

_____

97. For the days you had fever, for how long were your daily duties disrupted? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

DIZZINESS

97. Please tell me the number of days you had dizziness in the last two weeks. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

98. For the days you had dizziness, **at its worst**, how would you describe its intensity? ***[Tick one box only]***

1 = mild	2 = moderate	3 = severe	4 = very severe

99. Did dizziness disrupt your daily duties in any way?

1. Yes ***[go to 100].*** 2. No ***[go to 102].*** 3. DN***[go to 102].*** 4. NR ***[go to 102].***

100. How did dizziness disrupt your daily duties?

_____

_____

_____

101. For the days you had dizziness, for how long were your daily duties disrupted? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

[Go back to questions 36- 46 in Section on page 9 and say to the respondent.] **You have told me that you have had the following symptoms in the last two weeks *[read out the symptoms slowly].* Please think very carefully and tell me which of these symptoms bothered you most. *[Read out the symptoms again and list all those that bothered the respondent most].***

102. Which of the above symptoms would you say has bothered you most?
>>PROMPT: In what ways did this symptom bother you?<<

Symptoms	How did this symptom bother you?
1	
2	
3	
4	

103. Which symptoms would you say best represents illness with bilharzia (schistosomiasis) for you?

_____

104. In your own opinion and from your experience how can you best describe "life with bilharzia (schistosomiasis)"?

_____

105. In your own opinion how do you think having bilharzia can change or not change your health status and quality of life?

_____

106. Do you know of anyone who has suffered from bilharzia (schistosomiasis)? ***[Go to section C]***
1. Yes 2. No 3. DN 4. NR

SECTION C: EFFECTS ON HRQL DOMAINS

In this section I would like to ask you how bilharzia might or might not affect some areas of health related quality of life domains. In answering the questions, only consider your opinion and remember that there are no right or wrong answers. Please relate your responses to the following questions on your experience about illness with bilharzia.

[I will now read out some statements for you one by one. For each statements, please say whether you agree, disagree, do not know or not willing to respond]. Tick only one box against each question

		Agree	Disagree	DN	NR	
107	Being ill with bilharzia can affect one's mobility (for example walking about, going to workplace or being able to move about without any problems)	*[Go to 108]*	*[Go to 108]*	*[Go to 108]*	*[Go to 108]*	**Ask 113**
108	Being ill with bilharzia can affect one's performance of daily duties (e.g. playing, farming, household chores, usual activity [activity done for livelihood] schoolwork, etc)	*[Go to 109]*	*[Go to 109]*	*[Go to 109]*	*[Go to 109]*	*Ask 114*
109	Being ill with bilharzia can affect the performance and output of one's work (i.e. being able to accomplish as much as desired in activities of livelihood.)	*[Go to 110]*	*[Go to 110]*	*[Go to 110]*	*[Go to 110]*	*Ask 115*
100	Being ill with bilharzia can affect one's feeling of strength and energy in body	*[Go to 111]*	*[Go to 111]*	*[Go to 111]*	*[Go to 111]*	*Ask 116*
111	Being ill with bilharzia can affect one's ability to be in social functions such as group meetings, religious meetings, weddings, visiting friends and family and socializing with friends and colleagues.	*[Go to 112]*	*[Go to 112]*	*[Go to 112]*	*[Go to 112]*	*Ask 117*
112	Being ill with bilharzia can result in feelings of worry and anxiety.	*[Go to 108 last col.]*	*[Go to 108 last col.]*	*[Go to 108 last col.]*	*[Go to 108 last col.]*	*Ask 118*

[Go back to questions 107-112. For each question ticked "Agree", mark "X" in the last column and then ask the indicated questions on the column]

I would now like to ask you to tell me all the ways in which bilharzia affects each of the areas of health related quality of life that in your opinion you thought would be affected.

	How does bilharzia affect your	*[record all ways bilharzia affects these areas of health related quality of life*	
113	Mobility (for example walking about, going to workplace or being able to move about without problems)?		*Ask 119*
114	Performance of daily duties (e.g. playing, farming, household chores, usual activity [activity done for livelihood] schoolwork, etc)?		Ask 120
115	Performance and output of work (i.e. being able to accomplish as much as desired in activities of livelihood.)?		*Ask 121*
116	Feeling of strength and energy in body?		*Ask 122*
117	Ability to be in social functions such as group meetings, religious meetings, weddings, visiting friends and family and socializing with friends and colleagues?		*Ask 123*
118	Feelings of worry and anxiety?		*Ask 124*

[Go back to questions 113-118. For each question responded to tick "X" in the last column. Then ask the respondent to tell you how often the health related quality of life domain is affected and record in the table below]

I will now read out for you some health related quality of life domains that you said would be affected by due to illness with bilharzia. Please tell me how often each health related quality of life domain is affected.

Tick only one box against each health related quality of life domain.

		A little of the time	Some of the time	Most of the time
119	Please tell me how often your mobility is affected due to illness with bilharzia (for example walking about, going to workplace or being able to move about without problems).			
120	Please tell me how often your performance of daily duties is affected due to illness with bilharzia (e.g. playing, farming, household chores, usual activity [activity done for livelihood] schoolwork, etc)			
121	Please tell me how often performance and output of your work is affected due to illness with bilharzia (i.e. being able to accomplish as much as desired in activities of livelihood.).			
122	Please tell me how often your feeling of strength and energy in body is affected due to illness with bilharzia.			
123	Please tell me how often your ability to be in social functions such as group meetings, religious meetings, weddings, visiting friends and family and socializing with friends and colleagues is affected due to illness with bilharzia.			
124	Please tell me how often you feel worried and anxious due to illness with bilharzia.			

125. Are there any other ways you feel bilharzia (schistosomiasis) an affect your life or someone's life that you would like to tell me about?

_____

I will now read out for you some health related quality of life domains that you said would be affected by due to illness with bilharzia. Please tell me how often each health related quality of life domain is affected.

Tick only one box against each health related quality of life domain.

		A little of the time	Some of the time	Most of the time
119	Please tell me how often your mobility is affected due to illness with bilharzia (for example walking about, going to workplace or being able to move about without problems).			
120	Please tell me how often your performance of daily duties is affected due to illness with bilharzia (e.g. playing, farming, household chores, usual activity [activity done for livelihood] schoolwork, etc)			
121	Please tell me how often performance and output of your work is affected due to illness with bilharzia (i.e. being able to accomplish as much as desired in activities of livelihood.).			
122	Please tell me how often your feeling of strength and energy in body is affected due to illness with bilharzia.			
123	Please tell me how often your ability to be in social functions such as group meetings, religious meetings, weddings, visiting friends and family and socializing with friends and colleagues is affected due to illness with bilharzia.			
124	Please tell me how often you feel worried and anxious due to illness with bilharzia.			

125. Are there any other ways you feel bilharzia (schistosomiasis) an affect your life or someone's life that you would like to tell me about?

_____

SECTION D: EVALUATION QUESTIONS.

I will now go back to some of the questions I had asked you earlier and try to learn from you what meaning or understanding you attached to the questions and key words in those questions. Feel free and tell me as much as possible of what you think of these terms. Your opinion is what matters and there are no right or wrong answers.

126. Please describe in your own words what you consider to be **illness**.

_____

_____

127. Please describe in your own words what you consider to be **health problems**.

_____

_____

128. Please describe in your own words what you consider to be **daily duties**.

_____

_____

129. Please tell me in your own words what you consider **health state** to be.

_____

_____

SECTION E: GENERAL INFORMATION

To help in comparing responses from the various persons we have talked to I would like to ask you some general questions about yourself in this last section.

130 Name (*optional*)_____

131. Sex: 1. Male 2. Female *[circle one]*

132. What is your age? .....(*years*).

133. What is your current marital status? [*circle the correct response]*

1. Single 2. Married 3. Co-habiting
4. Divorced / Separated 5. Widower/ Widow 6. Other (specify)

134. What is the highest level of education that you have reached? *[Circle only one]*

	Educational level attained
1	None *[Do not ask 135]*
2	Primary
3	Secondary ("O" level)
4	Secondary ("A" level)
5	Diploma
6	Graduate (degree)
7	Adult education
8	Other (specify)

135.

How many years did you spend in *[Indicate number of years spent in school against each level up to the highest level reached as reported in 134 above]*	**Educational level attained**	**No. of years**
	Primary?	
	Secondary ("O" level)?	
	Secondary ("A" level)?	
	Diploma?	
	Graduate (degree)?	
	Adult education?	
	Other (specify)?	
Total years in school *[Do not add during interview. Add up after]*		

136. What is your main profession?

1 Farmer
2 Teacher
3 Businessperson (specify type)
4 Casual laborer
5 Civil servant (specify)
6 Student
7 Other (specify)

137. What do you do for a living?

1 Farmer
2 Teacher
3 Businessperson (Specify type)
4 Casual laborer
5 Civil servant (specify)
6 Student
7 Other (specify)

Record time interview ends. _____

Morning_____ Afternoon.

SECTION F: REQUEST TO PARTICIPATE IN THE NEXT PHASE OF THE STUDY

Thank you for answering all my questions. Before we finish, I would like to make a request from you. In the coming three to four months, we will do a study to establish the value or worth of the states of health resulting from the exercise we have just concluded with you. We would very much like you to participate in that exercise as a follow-up of this one. I am now kindly requesting you to tell me whether you are willing to be contacted for the follow up study.

138. Would you be willing to participate in the next phase of this study?

Yes *[Ask 139 and 140*	No *[End]*	DN *[End]*	NR *[End]*

139. Please tell me your physical address.
Name_____
Name of Village _____
Name of Village In-Charge _____

>>**PROMPT:** *Please tell me the easiest way to find you. When we get to the village from who or where can we easily trace your home?* **Take directions as exactly given**.

_____
_____
Nearest primary school_____ Nearest shopping center_____Nearest Church_____
Name of Sub-chief _____ Name of person by whom you can be easily found_____

140. Could you please tell me your postal contact address.
_____

We have now come to the end of this interview. I thank you very much for all you help and time. Is there anything you would like to ask me before we finish?

THANK YOU ONCE AGAIN FOR YOUR TIME AND ASSISTANCE.

SCHISTOSOMIASIS HRQL QUESTIONNAIRE

PARASITOLOGICAL AND CLINICAL ASSESSMENT

This clinical and parasitological information form will be first filled by the Research Assistant and assigned patient and questionnaire number at the same time as the HRQL questionnaire, following routine test and identification of schistosomiasis status, to facilitate matching later. As the LT conducts the Kato test, the patient will be sent back to the CO with this form for clinical examination and completion of section B. Upon completion, the patient will be sent back with this form to the Research Assistant who will administer the HRQL questionnaire. The Research Assistant will hand over this form to the Laboratory Technician for completion of section C and then match the form with the HRQL questionnaire. (*Section A should be filled in by the Research Assistant, Section B by the Clinical officer and section C by the Laboratory Technician)*

SECTION A: BACKGROUND INFORMATION ***(To be filled by Research Assistant and given to patient to take to Clinical Officer)***
[This questionnaire must be completed in pencil]

1	Questionnaire number		

		Day		Month		Year			
2	Date of interview.								

3. Study number and phase ***[Tick the appropriate box]***

1	2	3	4
S1P1 (Pre-test)	S1P2A (Patient study)	S1P2B (Community study)	S2 (Valuation study)

4.	Patient number		

5. Name of Patient _____

Recruited at	1	Kimbimbi Health Center
	2	Nguka Dispensary

6. Name of interviewer _____

7.

Location ***[Circle one only]***	1. Tebere
	2. Mutithi
	3. Thiba
	4. Nyangiti

	5. Murinduko

8. Sub-location ***[Circle one only]***

1	Kiarukungu	12	Wamumu
2	Mahigaini	13	Thiba
3	Gathigiriri	14	Mathangauta
4	Rukanga	15	Kirimara
5	Kinyaga	16	Ndomba
6	Mathigaini	17	Nyangati
7	Kombuini	18	Riagicheru
8	Kathiga	19	Mugabaciura
9	Kiandegwa	20	Miuu
10	Kabiriri	21	Kamunyange
11	Nguka		

SECTION B: GENERAL AND CLINICAL COMPLAINTS

[To be filled in by the Clinical Officer and returned to Research Assistant]

9. Name of Clinical Officer in Charge _____

10. Please write a verbatim detailed report of patient's general complaints at the time of hospital visit.

_____

_____

_____

11. Please write the clinical officer's diagnosis of the patient's health complaint at the time of hospital visit.

_____

_____

_____

12. Please indicate presence or absence of the following clinical signs of schistosomiasis in the patient
[To the clinician: Please fill in after physical examination]

Code		Yes=1	No=2
12a	Hepatomegaly		
12b	Splenomegaly		
12c	Anemia		
12d	Oedema		
12e	Ascites		
12f	Hepatic coma		
12g	Pulmonary hypertension		
12h	Periportal fibrosis		

13. Please indicate the magnitude of each clinical sign (***as measured clinically***)

Code		
13a	Hepatomegaly ***(cm)***	
13b	Splenomegaly ***(cm)***	
13c	Anemia ***(presence or Hb)***	
13d	Oedema ***(presence)***	
13e	Ascites ***(presence)***	
13f	Hepatic coma ***(Presence)***	
13g	Pulmonary hypertension ***(Presence)***	
13h	Periportal fibrosis ***(Presence)***	

SECTION C: PARASITOLOGICAL ASSESMENT

(To be filled in by the Laboratory Technician and forwarded to the Research Assistant)

14. Name of Laboratory Technician in charge_____
15. Schistosoma mansoni eggs per slide _____
16. Eggs per gram of feaces _____

17. Other parasites identified and eggs per slide and gram of feaces (where applicable).

CODE	**Name of parasite**	**Present**	**Absent**	**Eggs per slide (where applicable)**	**Eggs per gram of feaces (where applicable)**
17a					
17b					
17c					
17d					
17e					

>>Use additional space if required<<

APPENDIX 5.4: SCHISTOSOMIASIS HRQL QUESTIONNAIRE (kikuyu version)
(Questionnaire for patients)

[This questionnaire must be completed in pencil]

1	Questionnaire number		

		Day		Month		Year			
2	Date of interview.								

3. Study number and phase ***[Tick the appropriate box]***

1	2	3	4
S1P1 (Pre-test)	S1P2A (Patient study)	S1P2B (Community study)	S2 (Valuation study)

4.	Patient number		

4a. Name of patient_____

Recruited at	1	Kimbimbi Health Center
	2	Nguka Dispensary

5. Name of interviewer_____

6.

Location *[Circle one only]*	1. Tebere
	2. Mutithi
	3. Thiba
	4. Nyangiti
	5. Murinduko

7. Sub-location *[Circle one only]*

1	Kiarukungu	12	Wamumu
2	Mahigaini	13	Thiba
3	Gathigiriri	14	Mathangauta
4	Rukanga	15	Kirimara
5	Kinyaga	16	Ndomba
6	Mathigaini	17	Nyangati
7	Kombuini	18	Riagicheru
8	Kathiga	19	Mugabaciura
9	Kiandegwa	20	Miuu
10	Kabiriri	21	Kamunyange
11	Nguka		

INTRODUCTION

Njitagwo (uga ritwa riaku). Ndirarutithania wira na Mercy Mugo murutwo wa utuiria kuma University ya Nairobi. Turenda kumenya kuma kuri wee, mawoni maku ma uria murimu wa mbilihacia uthukagia ugima wa mwiri na muturire mwega wa andu. Tuguthurite kunyitanira na ithui githomo-ini giki kuringana na umenyo waku wa murimu uyu. Mawoni maku ni mabata muno hari kumenya wega uria murimu uyu uthukagia kana ukagera ngero muturire mwega wa andu. Ndihaha umuti ngiuria uteithie waku unjokerie ciuria ikonainie na murimu wa bilihacia. No tuthii na mbere? Ni wega muno.

(If yes, present the information sheet and consent form and allow the respondent to read through. If unable to read, read out for her/him. Ask the respondent to sign the consent form, before continuing with the interview).

[Interview starts]
Record start time ..... **Morning**..... **Afternoon** .....*{Tick as appropriate]*

SECTION A: HISTORY OF ILLNESS

Ngwambiriria na gukuria ciuria itarainie na rugano rwaku rwa urwaru wa ica ikuhi uria ungikorwo uguthinitie. Ugicokia ciuria ici, wicirie uhoro wa kahinda ga ciumia icio igiri irathirire. ***[Circle the correct response.]***

8. Ciumia icio igiri irathirire ri, niukoretwo na ndwari kana murimu wa muthemba o wothe?
   1. Ii ***[go to9].*** 2. Ari /Aca ***[go to11].*** 3. Ndiuii***[go to11]*** 4. Gutiri macokio ***[go to11]***
9. Niwamenyire ndwari iyo yari iriku (Niwamenyire murimu ucio wari uriku)?
   1. Ii ***[go to10].*** 2. Ari /Aca ***[go to11].*** 3. Ndiuii***[go to11]*** 4. Gutiri macokio ***[go to11]***
10. Yari iriku (Wari uriku) ? _____
11. Ciumia icio igiri irathirire ri, niukoretwo wina mathina ma ugima wa mwiri o na mariku?
    1. Ii ***[go to12].*** 2. Ari /Aca ***[go to17].*** 3. Ndiuii***[go to17]*** 4. Gutiri macokio ***[go to17]***
12. Ni wamenyire mathina macio ma ugima wa mwiri mari mariku?
    1. Ii ***[go to13].*** 2. Ari /Aca ***[go to14].*** 3. Ndiuii***[go to14]*** 4. Gutiri macokio ***[go to14]***

13. Mari mariku? _____

>> **Prompt:** *Ni kwari undu ungi?**[record all the illnesses and health problems the respondent mentions, each on a separate line. Use additional space below if needed]***

1. _____
2. _____
3. _____

14. Murimu ucio na mathina ma ugima wa mwiri ri, nimakugiririririe gwika maundu maku maria wikaga o muthenya o muthenya ?
1. Ii ***[Ask 15 and 16].*** 2. Ari /Aca ***[go to17].*** 3. Ndiuii***[go to17]*** 4. Gutiri macokio ***[go to17]***

15. No unjire ni maundu mariku maria wikaga o muthenya o muthenya murimu ucio wakugiririririe gwika.***[list each activity on a separate row. Use additional space if required]***

	Activity restricted
15a	
15b	
15c	
15d	
15e	

16. Hari o undu o undu wa maundu maya wanjira ri, wagiriririrrio kuwika ta mithenya iigana uguo. ***[Go back to responses in Q15 above and fill in the extent of restriction for each]***

Activity restricted	< 1 day	1-3 days	4-6 days	1-2 weeks
16a				
16b				
16c				
16d				
16e				

17. Riu ndagika ino-ri wi murwaru?
1. Ii ***[go to18].*** 2. Ari /Aca ***[go to20].*** 3. Ndiuii***[go to20]*** 4. Gutiri macokio ***[go to20]***

18. Ni uramenya kuria urwarite?
1. Ii ***[go to19].*** 2. Ari /Aca ***[go to20].*** 3. Ndiuii***[go to20]*** 4. Gutiri macokio ***[go to20]***

20. Ni ku? _____

20. Riu ndagika ino ri, wina thina wa ugima wa mwiri o wothe?
1. Ii ***[go to21].*** 2. Ari /Aca ***[go to23].*** 3. Ndiuii***[go to23]*** 4. Gutiri macokio ***[go to23]***

21. Ni uramenya ni mathina mariku ma ugima wa mwiri wina namo riu ndagika ino?
1. Ii *[go to22].* 2. Ari /Aca *[go to23].* 3. Ndiuii*[go to23]* 4. Gutiri macokio *[go to23]*

22. Ni mariku?

_____

23. Ni uri warwara murimu wa mbilihacia?
1. Ii *[go to24].* 2. Ari /Aca *[go to34].* 3. Ndiuii*[go to34]* 4. Gutiri macokio *[go to34]*

24. Wirigite kurwara mbilihacia ri?
Mweri__ __ Mwaka __ __ __ __ *[Go to 25]*

25. Niwacaririe uteithio wa njira ya guthodekwo?
1. Ii *[go to26].* 2. Ari /Aca *[go to33].* 3. Ndiuii*[go to33]* 4. Gutiri macokio *[go to33]*

26	Niwacaririe uteithio thibitari cia thirikari?	Ii *[go to 27]*	Ari/aca *[go to28]*	Ndiuii *[go to 28]*	Gutiri macokio *[go to28]*

27.Wacaririe uteithio thibitari cia thirikari iriku? ***[Circle all the facilities visited and indicate the name of the facility]*** ***>>PROMPT:*** *Ask the respondent to tell you the order in which they went to different facilities and rank as 1,2 3, ... from 1st to the last.*

code	Facility	Ranking of facility by order of visit
1	Kenyatta National Hospital	
2	Provincial Hospital	
3	District hospital	
4	Health Center	
5	Dispensary	

28	Ni wacaririe uteithio thibitari itari cia thirikari?	Ii *[go to 29]*	Ari/aca *[go to30]*	Ndiuii *[go to 30]*	Gutiri macokio *[go to30]*

29.Wacaririe uteithio thibitari iriku itari cia thirikari? ***[Circle all the facilities visited and indicate the name of the facility]*** *>> Ask the respondent to tell you the order in which they went to different facilities and rank as 1,2 3, ... from 1st to the last.*

code	Facility	Ranking of facility by order of visit
1	Private hospital	
2	Mission Hospital	
3	Mission dispensary	
4	Nursing home	
5	Private clinic	

30	Niwacaririe uteithio kuri andu aria mathodekananga gigitene?	Ii *[go to31]*	Ari/aca *[go to33]*	Ndiuii *[go to33]*	Gutiri macokio *[go to33]*
31	Niwacaririe uteithio kuri ndagitari cia miti?	Ii *[go to 32]*	Ari/aca *[go to33]*	Ndiuii *[go to33]*	Gutiri macokio *[go to33]*
32	Ni kuri kundu kungi wacaririe uteithio? (ku)	Ii *[go to 34]*	Ari/aca *[go to33]*	Ndiuii *[go to33]*	Gutiri macokio *[go to33]*

33. Waikarire atia na murimu ucio? _____

34. No unjirc uria ureigua ugima wa mwiri waku utarie ndagika ino

_____

>>***PROMPT:*** *Muhanire wa ugima wa mwiri uguiciria ni ta kii uguo?* ***[record the responses below]***

_____

_____

35. Tungihe mundu utari na thina wa ugima wa mwiri o na uriku makithi 100 rimwe, na mundu ukuite tumehe 0 ri, ungihe ugima wa mwiri waku makithi cigagana ndagika ino ? _____

Mundu ukuite — Mundu utari na thina wa ugima wa mwiri

PROP: 0 ——— 100

SECTION B: SYMPTOMS (CIONEKITHANIA CIA MIRIMU)

Riu ngukuria ciuria itarainie na cionekithania cia mirimu iria ungikorwo uiguite mwiri ini waku hari ciumia icio igiri ithirite

Nguguthomera cionekithania, nawe unjire imwe ciacio iria ungikorwo uiiguite mwiri-ini waku ciumia icio igiri ithirite.

[Tick one box only against yes, no, NA (not applicable) and NR (none response)]

		YES	NO	NA	NR	
36	Ciumia icio igiri ithirite ri, niukoretwo ukirio ni nda na kuigua nda uru?	*[Go to 37]*	*[Go to 37]*	*[Go to 37]*	*[Go to 37]*	**Ask 47-51**
37	>> PROMPT: riu ngukuria uhoro wa kuharwo. Ni kuri ruharo rutheri, nikuri ruharo ruhana mai, na nikuri ruharo ruri na thakame na rurenda. Tageria kunjira ruharo ruria ukoretwo naruo ciumia icio igiri ithirite<< Tick the diarrhea that fits the description as appropriate under Q37-39. Ciumia icio igiri ithirite ri, niukoretwo na ruharo rutheri?	*[Go to 38]*	*[Go to 38]*	*[Go to 38]*	*[Go to 38]*	**Ask 52-56**
38	Ciumia icio igiri ithirite ri, niukoretwo na ruharo ruhana mai?	*[Go to 39]*	*[Go to 39]*	*[Go to 39]*	*[Go to 39]*	**Ask 57-61**
39	Ciumia icio igiri ithirite ri, niukoretwo na ruharo rwina thakame na rurenda?	*[Go to 40]*	*[Go to 40]*	*[Go to 40]*	*[Go to 40]*	**Ask 62-66**
40	Ciumia icio igiri ithirite ri, niwaiguaga kunogerera?	*[Go to 41]*	*[Go to 41]*	*[Go to 41]*	*[Go to 41]*	**Ask 67-71**
41	Ciumia icio igiri ithirite ri, niwaiguauga kuira ngoro?	*[Go to 42]*	*[Go to 42]*	*[Go to 42]*	*[Go to 42]*	**Ask 72-76**
42	Ciumia icio igiri ithirite ri, niutahikite thakame?	*[Go to 43]*	*[Go to 43]*	*[Go to 43]*	*[Go to 43]*	**Ask 77-81**
43	Ciumia icio igiri ithirite ri, niuiguite utakwenda kuria kana kunyua?	*[Go to 44]*	*[Go to 44]*	*[Go to 44]*	*[Go to 44]*	**Ask 82-86**
44	Ciumia icio igiri ithirite ri, niukoretwo na tuthundo twina mwithua ngothi-ini?	*[Go to 45]*	*[Go to 45]*	*[Go to 45]*	*[Go to 45]*	**Ask 87-91**
45	Ciumia icio igiri ithirite ri, niuiguite kuhiuha mwiri nakuigua kiurugari?	*[Go to 46]*	*[Go to 46]*	*[Go to 46]*	*[Go to 46]*	**Ask 92-96**
46	Ciumia icio igiri ithirite ri, niuiguite thiurura?	*[Go to 102]*	*[Go to 104]*	*[Go to 104]*	*[Go to 104]*	**Ask97-102**

[Go through the column marked "yes" in questions 36-47 above and for each "yes" answer, tick "X" on the last column. For each box ticked "X" ask the questions indicated]

ABDOMINAL PAIN AND DISCOMFORT (KURIO NI NDA NA KUIIGUA URU NDA)

47. Ugire niwariagwo ni nda na kuigua nda uru ciumia icio igiri ithirite. No unjire ni ta mithenya iigana uguo wariirwo ni nda na kuigua nda uru .

[Tick one box only]

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

48 Hari mithenya iyo wariagwo ni nda na kuigua nda uru ri, mithenya iria waiguaga uru muno ri, ruo rwatarii atia? ***[Tick one box only]***

1 = runini	2 = ruingaingi	3 = ruingi	4 = ruingi muno

49. Ruo ruu rwa nda na kuigua uru nda nikwarigagiriria gwika maundu maku maria wikaga o muthenya o muthenya?

1. Ii ***[go to 50].*** 2. Ari/Aca ***[see 37 last col.].*** 3. Ndiuii ***[see 37 last col.]*** 4. Gutiri macokio***[see 37 last col.]***

50. Kwarigagiriria atia?

_____

_____

_____

51. Ni ta mithenya iigana uguo warigiriirio gwika maundu macio ma o muthenya o muthenya? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

DIARRHEA (KUHARWO GUTHERI)

52. Ni mithenya iigana waharirwo ruharo rutheri ciumia icio igiri ithirite. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

53 Hari mithenya iyo ri, mithenya iria ruharo rwari ruru muno ri, waharirwo ta maita maigana hari muthenya?***[Tick one box only]***

1 = 3-9	2 = 9-15	3 = 15-21	4 = over 21

54. Ni rwa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya?

1. Ii ***[go to 55].*** 2. Ari/Aca ***[see 38 last col.].*** 3. Ndiuii ***[see 38 last col.]*** 4. Gutiri macokio***[see 38 last col.]***

55. Rwa kurigiriirie atia?

_____

_____

_____

56. Ni ta mithenya iigana uguo rwa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

WATERY DIARRHEA (RUHARO RWA NA MAI)

58. Ni mithenya iigana waharirwo ruharo rwi na mai ciumia icio igiri ithirite. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

58. Hari mithenya iria rwari ruru muno ri, waharirwo ta maita maigana uguo hari mutheya? ***[Tick one box only]***

1 = 3-9	2 = 9-15	3 = 15-21	4 = over 21

59. Ni rwa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya?

1. Ii ***[go to 60].*** 2. Ari/Aca ***[see 39 last col.].*** 3. Ndiuii ***[see 39 last col.]*** 4. Gutiri macokio***[see 39 last col.]***

60. Rwa kurigiriirie atia?

_____

_____

_____

62. Ni ta mithenya iigana uguo rwa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

BLOODY MUCOID DIARRHEA (RUHARO RWINA THAKAME NA RURENDA)

62. Ni mithenya iigana uguo waharirwo ruharo rwina thakame na rurenda ciumia icio igiri ithirite? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

63 Hari mithenya iria rwari ruru muno ri, waharirwo maita ta maigana uguo hari muthenya? ***[Tick one box only]***

1 = 3-9	2 = 9-15	3 = 15-21	4 = over 21

64. Ni rwa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya?

1. Ii ***[go to 65].*** 2. Ari/Aca ***[see 40 last col.].*** 3. Ndiuii ***[see 40 last col.]*** 4. Gutiri macokio***[see 40 last col.]***

65. Rwa kurigiriirie atia?

_____

_____

_____

66. Ni ta mithenya iigana uguo rwa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

TIREDNESS (KUNOGERERA)

67. Ni mithenya iigana ugou waiguaga kunogerera ciumia icio igiri ithirite. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

68. Hari mithenya iria waiguaga unogereire muno ri, waiguaga minoga iyo itarii ta atia? ***[Tick one box only]***

1 = kunoga hanini	2 = kunogerera	3 = kunoga muno	4 = kunoga muno makiria

69. Ni ya kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya?

1. Ii ***[go to 70].*** 2. Ari/Aca ***[see 41 last col.].*** 3. Ndiuii ***[see 41 last col.]*** 4. Gutiri macokio***[see 41 last col.]***

70. Ya kurigiriirie atia?

_____

_____

71. Ni ta mithenya iigana uguo yakurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

NAUSEA (KUIRA NGORO)

72. Ni mithenya iigana waiguaga kuira ngoro ciumia icio igiri ithirite?. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

73. Hari mithenya iria waiguaga kuira ngoro muno ri, kiiria kia ngoro giatarii atia? ***[Tick one box only]***

1 = kinini	2 = kiingaingi	3 = kiingi	4 = kiingi muno

74. Ni gia kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya?

1. Ii ***[go to 75].*** 2. Ari/Aca ***[see 42 last col.].*** 3. Ndiuii ***[see 42 last col.]*** 4. Gutiri macokio***[see 42 last col.]***

75. Gia kurigiriirie atia?

_____

_____

_____

76. Ni ta mithenya iigana uguo gia kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya ? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

VOMITING BLOOD (GUTAHIKA THAKAME)

77. Ni mithenya iigana uguo watahikire thakame hari ciumia icio igiri ithirite. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

78. Hari mithenya iria rutahiko rwari ruru muno ri, ungitariria uria thakame iria yari matahiko ini yaiganaga atia? ***[Tick one box only]***

1 = Nini muno	2 = nini	3 = nyingaingi	4 = nyingi muno

79. Ni rwa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya?

1. Ii ***[go to 80].*** 2. Ari/Aca ***[see 43 last col.].*** 3. Ndiuii ***[see 43 last col.]*** 4. Gutiri macokio***[see43 last col.]***

80. Ra kurigiriirie atia?

_____

_____

_____

81. Ni ta mithenya iigana uguo rwa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

LOSS OF APPETITE (KUIGUA UTAKWENDA KURIA KANA KUNYUA)

82. Ni ta mithenya iigana uguo waiguaga utakwenda kuria kana kunyua hari ciumia icio igiri ithirite. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

83. Hari mithenya iria waiguaga utakwenda kuria kana kunyua muno rı, uringithanitie na muigana wa irio iria uriaga kana kiria unyuaga riria utakuigua uguo rı, ungiuga wahotire kuria kana kunyua muigana uriku? ***[Tick one box only]***

1 = Ta kuota ithatu uguo	2 = ta nuthu kinya kuota imwe	3 = ta iciko igiri uguo ngaremwo	4= ndingiria kana nyue o na ki

84. Kuremwo ni kuria ni gwa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya?

1. Ii ***[go to 85].*** 2. Arı/Aca ***[see 44 last col.].*** 3. Ndiuii ***[see 44 last col.]*** 4. Gutiri macokıo***[see 44 last col.]***

86. Gwa kurigiriirie atia?

_____

_____

_____

86. Ni ta mithenya iigana uguo gwa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

ITCHING SKIN RASH (TUTHUNDO TWI NA MWITHUA NGOTHI-INI)

87. Ni ta mithenya iigana uguo wari na tuthundo twi na mwithua ngothi-ini ciumia icio igiri ithirite. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

88. Hari mithenya iria mwithua wa tuthundo tuu wari muru muno ri, ungitariria watarii ta atia? ***[Tick one box only]***

1 = munini	2 = muingaingi	3 = muingi	4 = muingi muno

89. Ni wa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya?

1. Ii ***[go to 90].*** 2. Arı/Aca ***[see 45 last col.].*** 3. Ndiuii ***[see 45 last col.]*** 4. Gutiri macokio***[see 45 last col.]***

90. Wakurigiriirie atia?

_____

_____

_____

91. Ni ta mithenya iigana uguo wa kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

FEVER (KUHIUHA MWIRI NA KIURUGARI)

92. Ni mithenya iigana uguo waiguaga kuhiuha mwiri na kiurugari hari ciumia icio igiri ithirite?. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

93. Hari mithenya iria kuhiuha mwiri na kiurugari kiari kiuru muno ri, ungitariria giatarie atia? ***[Tick one box only]***

1 = Kinini	2 = kiingaingi	3 = kiingi	4 = kiingi muno

94. Ni gia kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya?

1. Ii ***[go to 95].*** 2. Ari/Aca ***[see 46 last col.].*** 3. Ndiuii ***[see 46 last col.]*** 4. Gutiri macokio***[see 46 last col.]***

95. Gia kurigiriirie atia?

_____

_____

97. Nita mithenya iigana uguo gia kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

DIZZINESS (THIURURA)

97. Ni ta mithenya iigana uguo waiguaga thiurura hari ciumia icio igiri ithirite?. ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

98. Hari mithenya iria thiurura yari njuru muno ri, ungitariria yatarii atia? ***[Tick one box only]***

1 = nini	2 = nyingaingi	3 = nyingi	4 = nyingi muno

99. Ni ya kurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya?

1. Ii ***[go to 100].*** 2. Ari/Aca ***[Go to 102].*** 3. Ndiuii ***[Go to 102]*** 4. Gutiri macokio***[Go to 102]***

100. Ya kurigiriirie atia?

_____

_____

_____

101. Ni ta mithenya iigana uguo yakurigiriirie gwika maundu maku maria wikaga o muthenya o muthenya? ***[Tick one box only]***

1= Less than a day	2= 1-3 days	3= 4-6 days	4= 7-10 days	5= 11-14 days

[Go back to questions 36- 46 in Section B on page 9 and say to the respondent.] **Unjirire ciumia icio igiri ithirite wari na cionekithania ici cia mirimu *[read out the symptoms slowly].* Riu ngwenda unjire ni iriku ciaicio iguthumburite muno *[Read out the symptoms again and list all those that bothered the respondent most].***

102. Ni cionekithania iriku cia mirimu iraguthumburire muno makiria ciumia icio igiri irathirire?

>>PROMPT: In what ways did this symptom bother you?<<

Cionekithania cia mirimu	Kionekithania giki giguthumburite atia?
1	
2	
3	
4	
5	
6	
7	
8	

103. Hari cionekithania ici ri, ni iriku ciacio ungiuga cionanagi ati mundu ari na mbilihacia wega?

104. Kuringana na woni waku na umenyo waku ri, ungitariria muturire wa mundu ena mbilihacia ukoragwo utarie atia?

105. Kuringana na woni waku ri, gokorwo wina mbilihacia kungigarurira kana kwage kugarurira ugima wa mwiri na muturire mwega waku atia ?

106. Ni uii mundu uri warwara mbilihacia ? ***[Go to section C]***

1. Ii 2. Ari/aca 3. Ndiuii 4. Gutiri macokio

SECTION C: EFFECTS ON HRQL DOMAINS

Riu ngukuria ciuria ikonainie no uria mbilihacia ingithukia kana yage guthukia miena imwe na imwe ya ugima wa mwiri ikonainie na muturire mwega wa mundu. Ugicokia ciuria ici, ngwenda uhe mawoni maku, ukimenyaga gutiri macokio ati magiririire kana matagiriire. Macokio o mothe ni mega. Ningi ugicokia ciuria ici ringithania macokio maku na uria uuii ukonainie na murimu wa mbilihacia.

Riu nguguthomera maundu me haha umwe kwa umwe. Nawe o hari undu unjire kana niwetikaniria na undu ucio kana niwaregana naguo, kana nduramenya.

Tick only one box against each question.

		Agree	Disagree	DN	NR	
107	Gukorwo na mbilihacia no gutume mundu aremwo ni gucera cera na guthianga (ta kuhota guthii, guthii wira-ini, kana guthianga utari na gathina o na kamwe)	*[Go to 108]*	*[Go to 108]*	*[Go to 108]*	*[Go to 108]*	Ask 113
108	Gukorwo na mbilihacia no guthukie mwikire wa maundu maria mundu ekaga o muthenya o muthenya (ta mithako, kurima, mawira ma nyumba, wira uria mundu arutaga kuhingia mabataro make, wira wa cukuru, etc)	*[Go to 109]*	*[Go to 109]*	*[Go to 109]*	*[Go to 109]*	*Ask 114*
109	Gukorwo na mbilihacia no guthukie mwikire na maciaro ma wira wa mundu (ta gwika maundu maria mothe ungieenda gwika kwiethera uturo)	*[Go to 110]*	*[Go to 110]*	*[Go to 110]*	*[Go to 110]*	*Ask 115*
100	Gukorwo na mbilihacia no gutume mundu aiigue atari na uhoti na hinya mwiri ini.	*[Go to 111]*	*[Go to 111]*	*[Go to 111]*	*[Go to 111]*	*Ask 116*
111	Gukorwo na mbilihacia nogutume mundu aremwo ni gutukana na kunyitanira na andu (ta guthii micemanio ya ikundi, ya kanitha, mohiki, gucerera arata na andu a nyumba na gukenanira na arata na aria murutithanagia wira nao).	*[Go to 112]*	*[Go to 112]*	*[Go to 112]*	*[Go to 112]*	*Ask 117*
112	Gukorwo na mbilihaciano gutume mundu aiigue ena kimako na gitangiko kiingi.	*[Go to 108 last col.]*	*[Go to 108 last col.]*	*[Go to 108 last col.]*	*[Go to 108 last col.]*	*Ask 118*

[Go back to questions 107-112. For each question ticked "Agree", mark "X" in the last column and then ask the indicated questions on the column]

Riu no nyende unjire uria mbilihacia ithukagia o umwe wa miena imwe na imwe ya ugima wa mwiri ikonainie na muturire mwega wa mundu kuringana na mawoni maku makonie miena iria wiciririe no ithukio.

	Uguiciria mbilihacia ingithukia	*[record all ways bilharzia affects these areas of health related quality of life*	
113	Gucera cera na guthianga (ta kuhota guthii, guthii wira ini kana guthianga utari na gathina o na kamwe) atia?		*Ask 119*
114	Mwikire wa maundu maria mundu ekaga o muthenya o muthenya (ta mithako, kurima, mawira ma nyyumba, wira uria mundu arutaga kuhingia mabataro make, wira wa cukuru etc) atia?		Ask 120
115	Mwikire na maciaro ma wira wa mundu (ta gwika maundu maria mothe ungienda gwika kwiethera uturo) atia?		*Ask 121*
116	Ithukagia uhoti na hinya thiini wa mwiri atia?		*Ask 122*
117	Ugwiciria mbilihacia ingiremithia mundu gutukana na kunyitanira na andu (ta guthii micemanio ya ikundi, ya kanitha, mohiki, gucerera arata na andu a nyumba na gukenanira na arata na aria murutithanagia wira nao) atia?.		*Ask 123*
118	Ugwiciria mbilihacia itumaga mundu aiigue ena kimako na gitangiko niki (kana atia)?		*Ask 124*

[Go back to questions 113-118. For each question responded to tick "X" in the last column. Then ask the respondent to tell you how often the health related quality of life domain is affected and record in the table below]

Riu nguguthomera miena imwe na imwe ya ugima wa mwiri ikonainie na muturire mwega wa mundu, iria wanjira mbilihacia no ithukie. O hari mwena ri, ta njira mbilihacia ingithukia ta kahinda kaigana atia?

Tick only one box against each health related quality of life domain.

		Kahinda kanini	**Rimwe na rimwe**	**Mahinda maingi**
119	Kuremwo ni gucera cera na guthianga (ta kuhota guthii, guthii wira-ini, kana guthianga utari na gathina o na kamwe)			
120	Mwikire wa maundu maria mundu ekaga o muthenya o muthenya (ta mithako, kurima, mawira ma nyumba, wira uria mundu arutaga kuhingia mabataro make, wira wa cukuru, etc)			
121	Mwikire na maciaro ma wira wa mundu (ta gwika maundu maria mothe ungieenda gwika kwiethera uturo)			
122	Kuigie utari na uhoti na hinya mwiri ini.			
123	Kuremwo ni gutukana na kunyitanira na andu (ta guthii micemanio ya ikundi, ya kanitha, mohiki, gucerera arata na andu a nyumba na gukenanira na arata na aria murutithanagia wira nao).			
124	Kuigue wina kimako na gitangiko kiingi.			

125. Ni hari maundu mangi ukwona mbilihacia ithukagia muturire wa mundu ungienda kunjira?

_____

_____

SECTION D: EVALUATION QUESTIONS.

Riu ndirenda gukugwetera ciugo imwe ngugwetete hau kabere. Ndirenda kumenya uria unyitire ciugo icio cikugaga. Njira maundu maingi o uria kwahoteka o hari kiugo utakumaka kana uria uroiga niguo wagiriire kana tiguo wagiriire. Mawoni maku nimo mari na bata haha.

126. Ndariria wiciririe **ndwari** kana **murimu** ni kii?

_____

_____

127. Ndariria na ciuogo ciaku wiciririe **mathina ma ugima wa mwiri** ni kii?

_____

_____

128. Ndariria na ciugo ciaku **mawira maku maria wikaga o muthenya o muthenya** ni mariku?

_____

_____

129. Ndariria **ugima wa kimwiri** ni ta kii uguo?

_____

_____

SECTION E: GENERAL INFORMATION

Nigetha tuhote kuringithania macokio ma andu aria othe twariirie ri, ni ngukuria ciuria igukonie.

130 Ritwa (*optional*)_____

131. Kiumbe: 1. ndume 2. Nga *[circle one]*

132. Wina miaka iigana? .....(miaka).

133. Muikarire waku ni uriku? [*circle the correct response]*

1.Wike 2. Muhiku/kuhikania 3. muikaranio

4. Gutigana 5. Muthuri/mutumia wa ndigwa 6. ungi (uga)

134. Wakinyirie githomo giaku ha? *[Circle only one]*

	Educational level attained
1	Noti *[Do not ask 135]*
2	Primari
3	Secondari (fomu 4)
4	Secondari (fomu 6)
5	Diploma
6	Graduate (degree)
7	Ngubaru
8	Hangi (uga)

135.

	Educational level attained	**No. of years**
O hari levo ya githomao waikarire miaka iigana? *[Indicate number of years spent in school against each level up to the highest level reached as reported in 134 above]*	Primari	
	Secondari (fomu 4)	
	Secondari (fomu 6)	
	Diploma	
	Graduate (degree)	
	Ngubaru	
	Hangi (uga)	
Total years in school *[Do not add during interview. Add up after]*		

136. Uthomeire wira uriku?

1 Murimi
2 Mwarimu
3 Mubiachara (uga mbiachara)
4 wira wa moko
5 muruti wa wira wa thirikari (uga wira uriku)
6 murutwo
7 ungi (uga)

137. Wikaga atia wagukuhotithia kuhingia mabataro maku?

1 Murimi
2. Mwarimu
3 Mubiachara (uga mbiachara)
4 wira wa moko
5 muruti wa wira wa thirikari (uga wira uriku)
6 murutwo
7 ungi (uga)

Record time interview ends. _____ Morning

_____ Afternoon.

SECTION F: REQUEST TO PARTICIPATE IN THE NEXT PHASE OF THE STUDY

Ni wega muno ni kunjokeria ciuria. Tutanirikia nonyende gukuria uteithio. Mieri itatu kana ina yukite, nitukageria kumenya uria tungihota gutarania na kuhe miugana mithemba ya ugima wa mwiri iria ikumana na ciuria icio ndakuria umuthi. No twende muno ukanyitanira na ithui gutarania ini kuu tondu uri umwe wa aria mateithia kwonekithania mithemba iyo ya ugima wa kimwiri turenda gu gatara na kuhe muigana. Riu ndirenda gukuria kana no ngucerere ringi hindi iyo yakinya.

138. Ngoka gugucerera ringi?

Ii *[Ask 139 and 140*	Ari/aca *[End]*	Ndiuii *[End]*	Gutiri macokio *[End]*

139. Njira uria ingikwona.

Ritwa_____

Ritwa ria gicagi_____

Ritwa ria In-Charge wa gicagi _____

>>**PROMPT: Ndariria uria ingiuka kwanyu raithi. [Take directions as exactly given]**.

_____

_____

Primari iria ii hakuhi_____ shopping center _____Chachi iria ii hakuhi_____
Ritwa ria Sub-chief _____ Ritwa riria tungiuria gukwona narua_____

140. He namba ya ithanduku ria marua

_____

NI WEGA MUNO. NI NDACOKIA NGATHO O RINGI NIUNDU WA UTEITHIO UCIO WOTHE WAHEE. NGAI AKURATHIME.

APPENDIX 6.1: FIGURES

Figure A6.1: Frequency of symptoms in the last two weeks

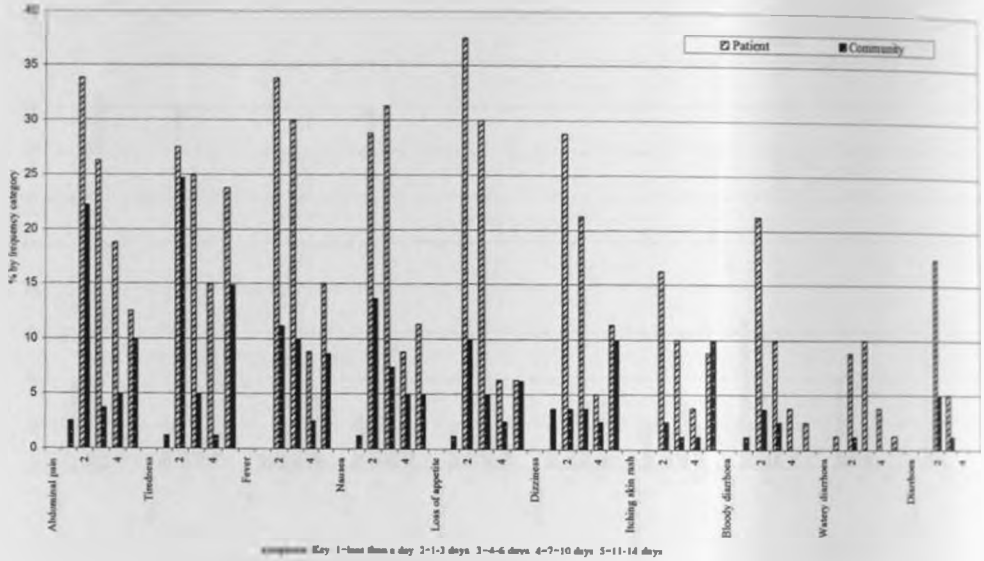


Figure A6.2: Intensity of symptoms in the last two weeks

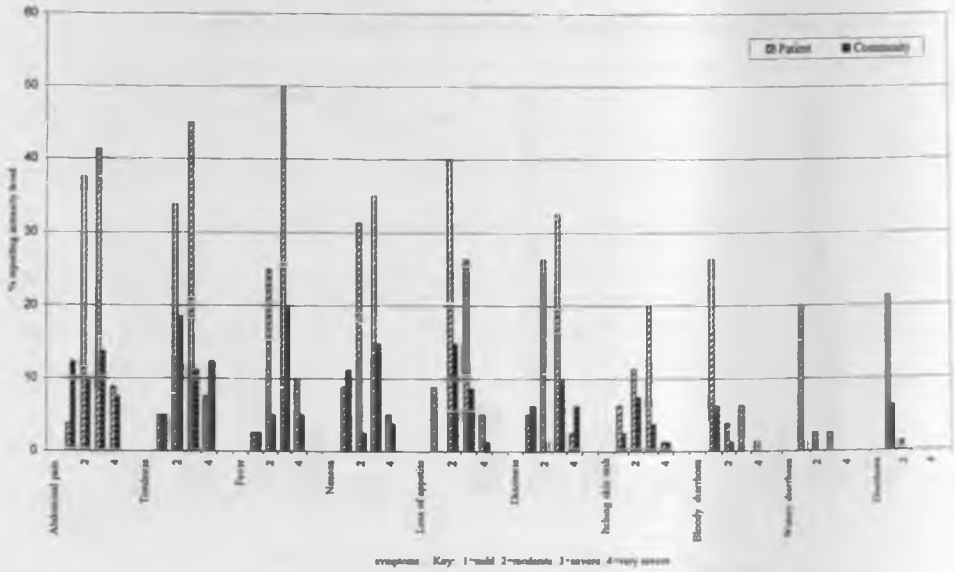


Figure A6.3: Symptom severity in the last two weeks

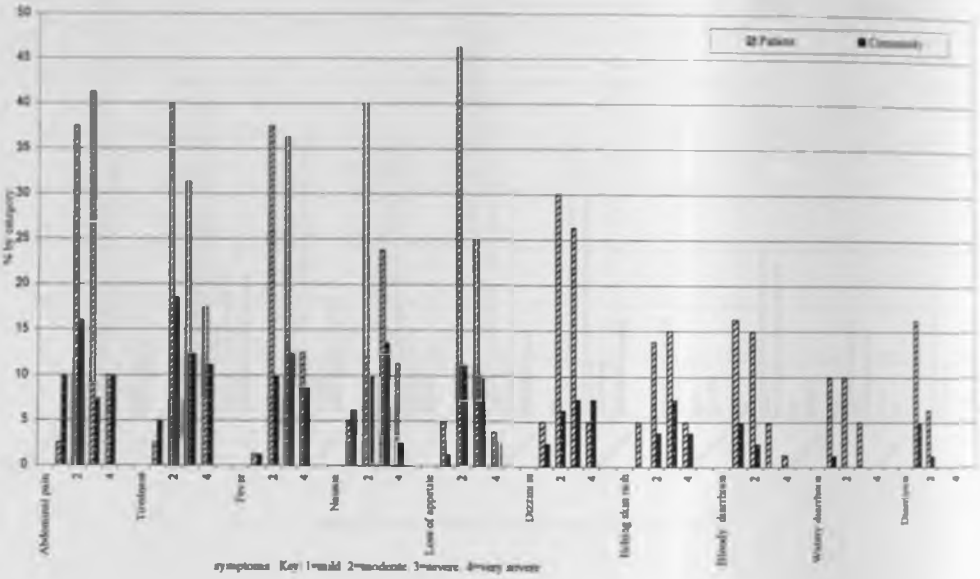


Figure A6.4: Overall symptom indices (frequency, intensity and severity) in the last two weeks

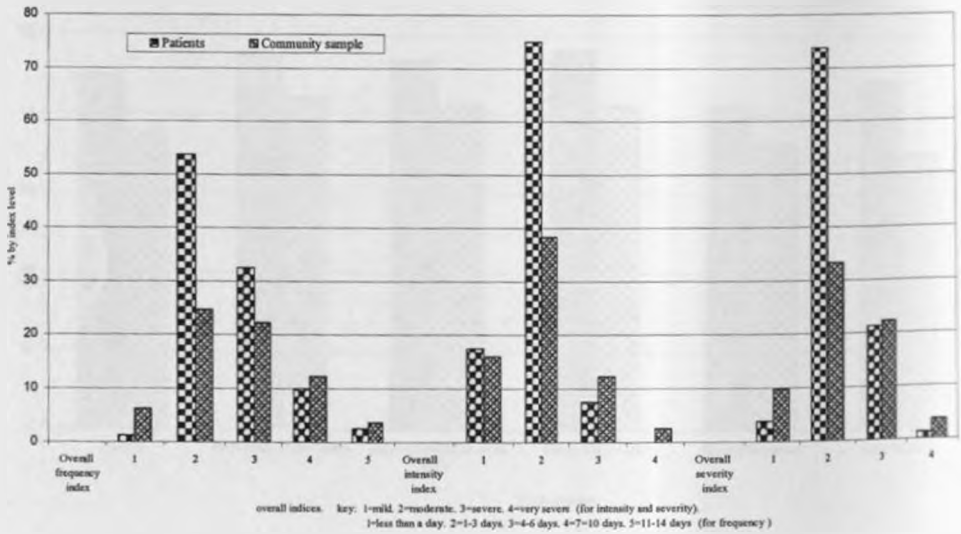


Figure A6.5: Frequency of disruption of daily duties in the last two weeks

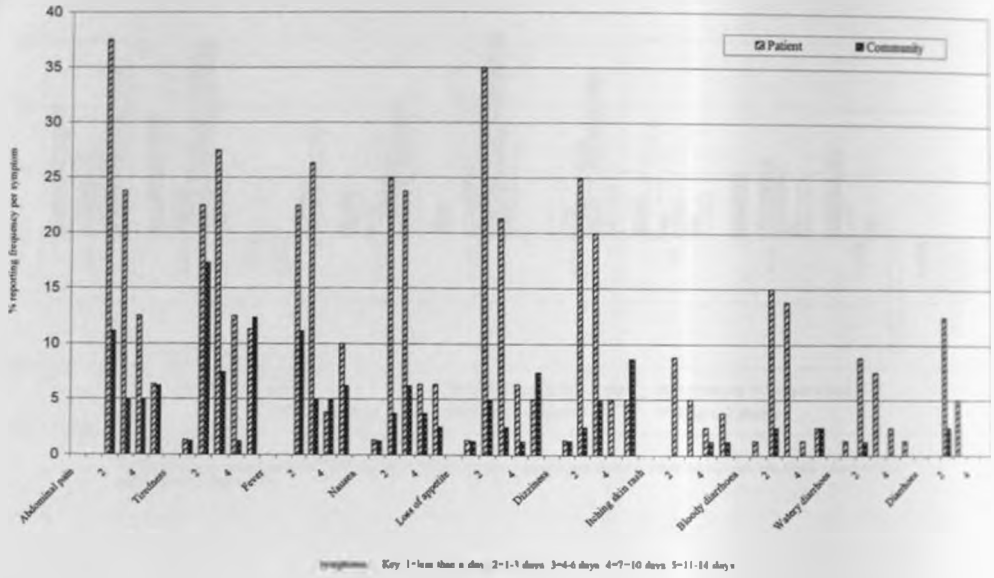


Figure A6.6: % of respondents agreeing that HRQL domains are affected by bilharzia.

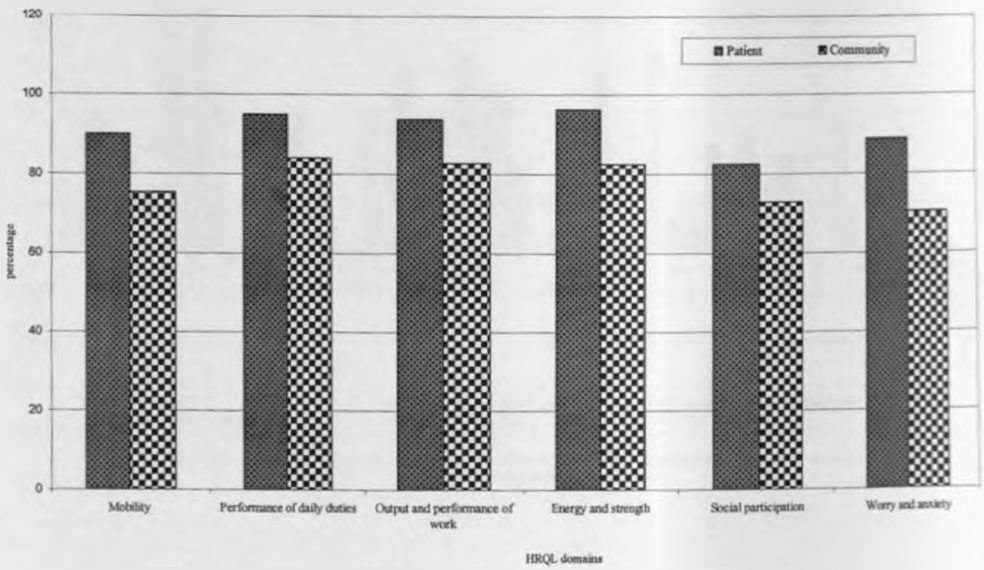
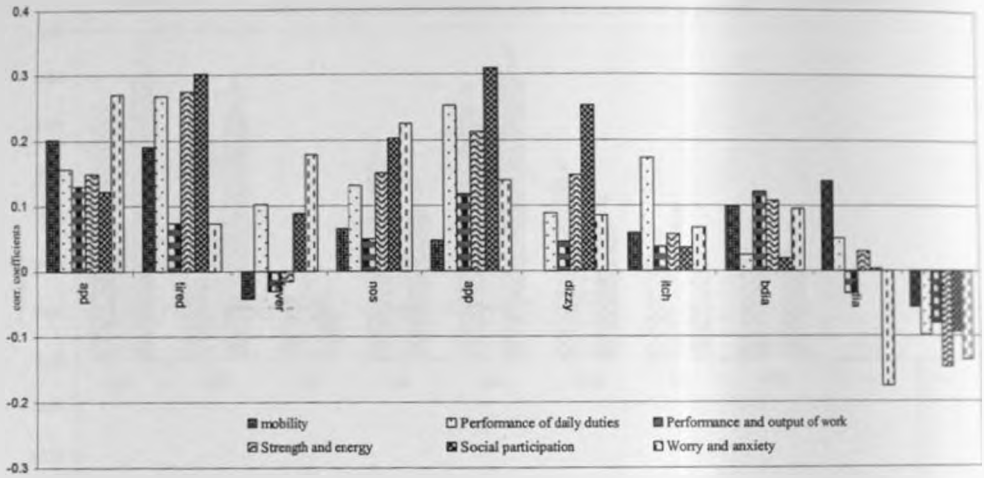
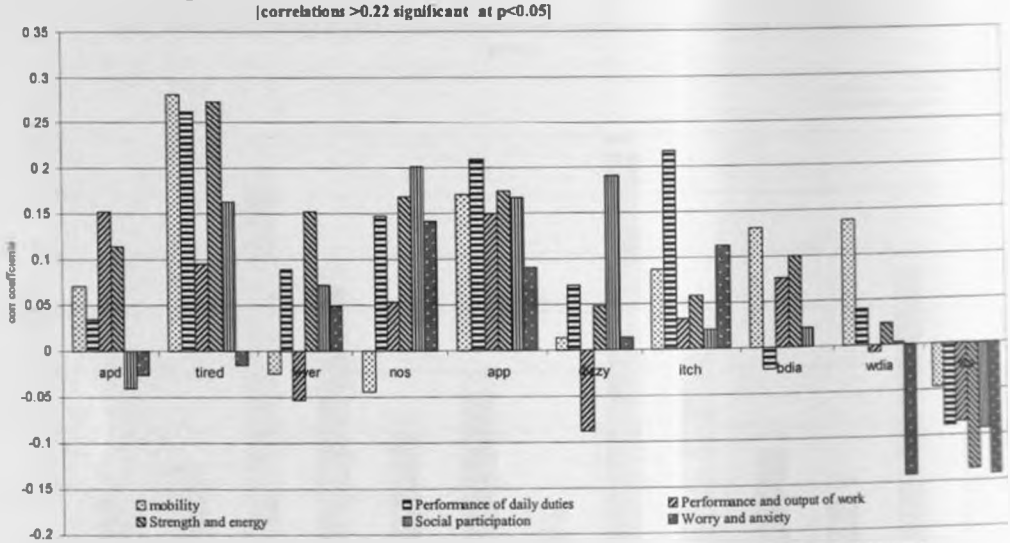


Figure A6.7: Correlations between symptom frequency and HRQL domains (patient group). |correlations >0.22 significant at p<0.05|



key apd-abdominal pain and discomfort Tired-tiredness, nos-nausea, app-loss of appetite, dizzy-dizziness, itch-itching skin rash, bdia-bloody diarrhoea, wdia-watery diarrhoea, din-diarrhoea

Figure A6.8: Correlations between intensity of symptoms and HRQLdomains (patient group) |correlations >0.22 significant at p<0.05|



symptoms key apd-abdominal pain and discomfort Tired-tiredness, nos-nausea, app-loss of appetite, dizzy-dizziness, itch-itching skin rash, bdia-bloody diarrhoea, wdia-watery diarrhoea, din-diarrhoea

Figure A6.9: Correlations between health status index and symptoms frequency, intensity and severity aggregate indices (patient group). [correlations >0.22 significant at p<0.05]

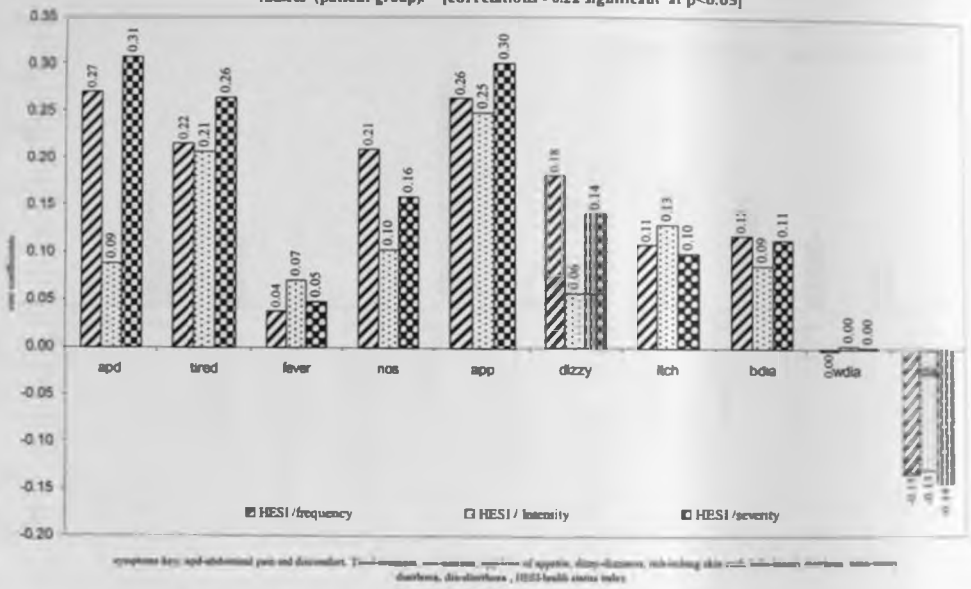


Figure A6.10: Kruskal Wallis P values for association between symptom severity and health status index (patient group)

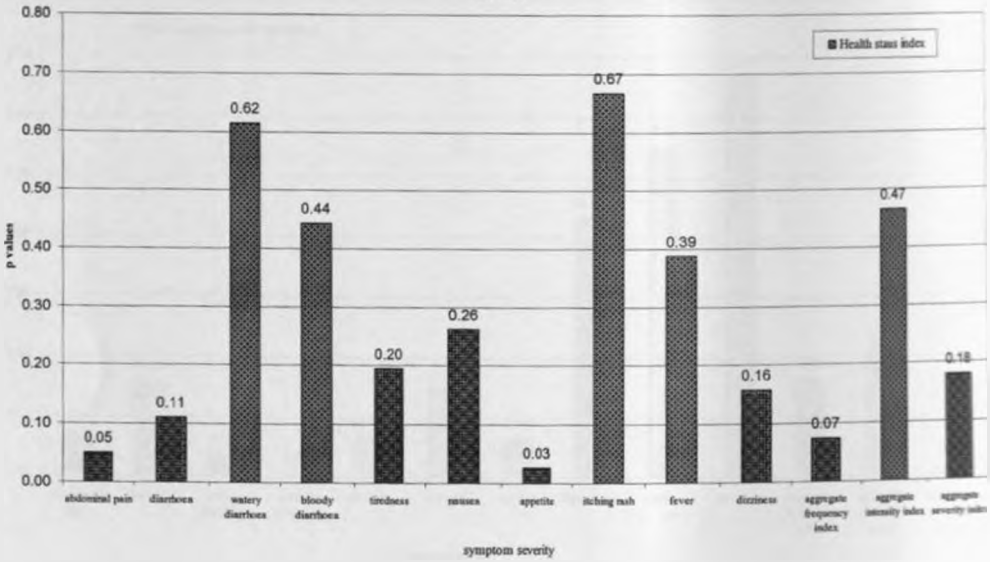


Figure A6.11: Correlation and *p*- values for VAS rating by symptom severity index (patient group)

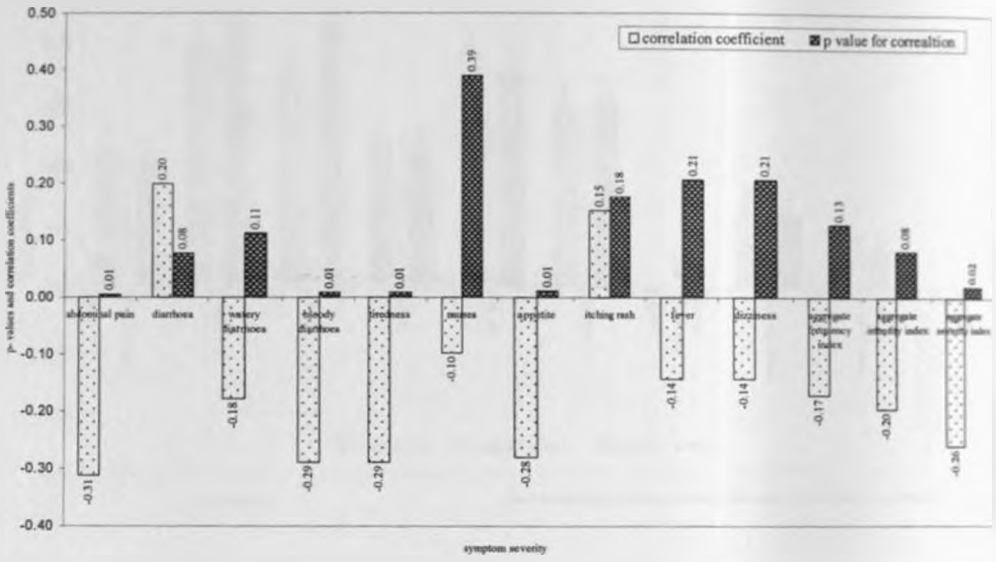


Figure A6.12: Kruskal Wallis p values for VAS rating of current health state and symptom severity (patient group)

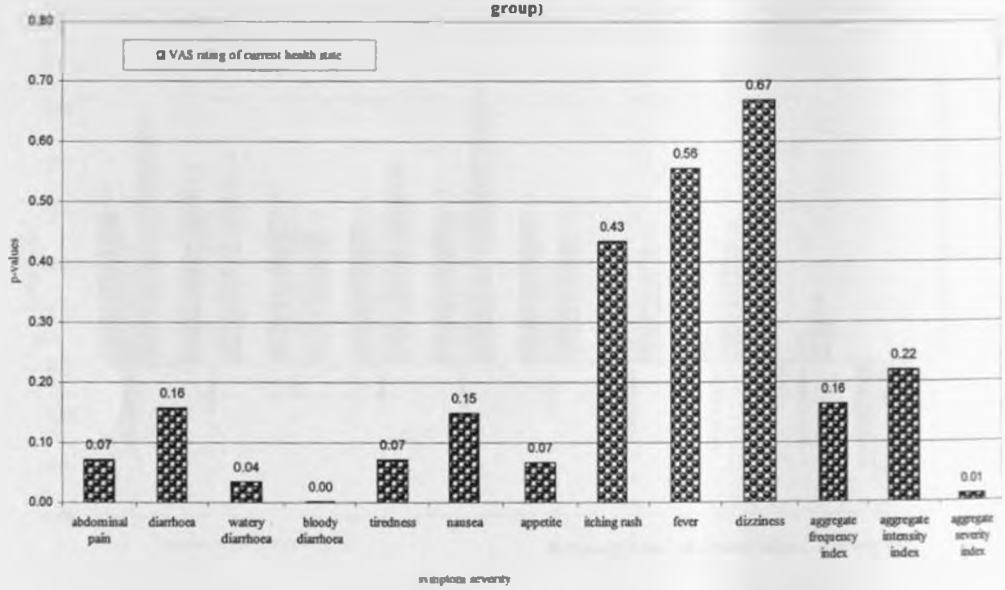


Figure A6.13: Correlations between frequency, intensity and severity of symptoms and frequency of disruption of daily duties (patient group) [correlations >0.22 significant at p<0.05]

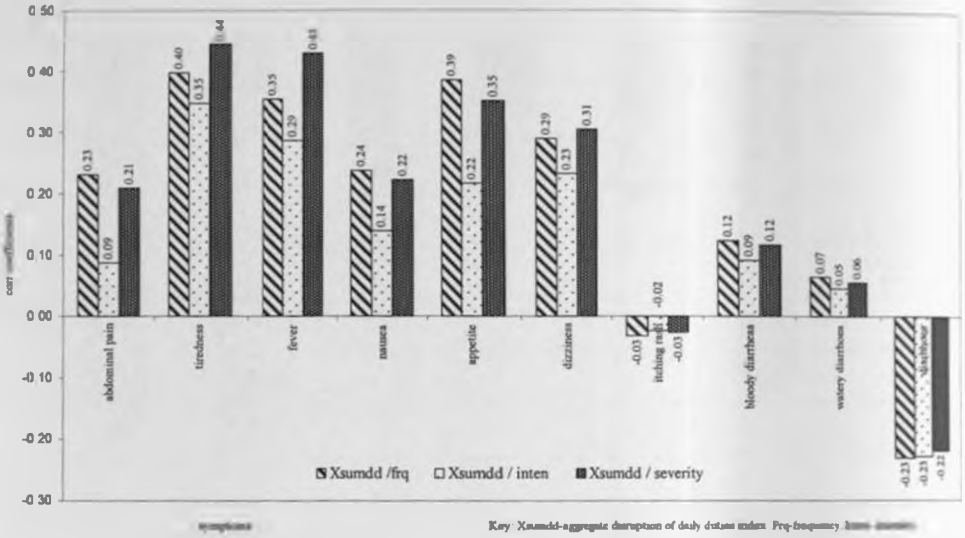


Figure A6.14: Correlations between aggregate symptom indices and frequency of disruption of daily duties (patient group) [correlations >0.22 significant at p<0.05]

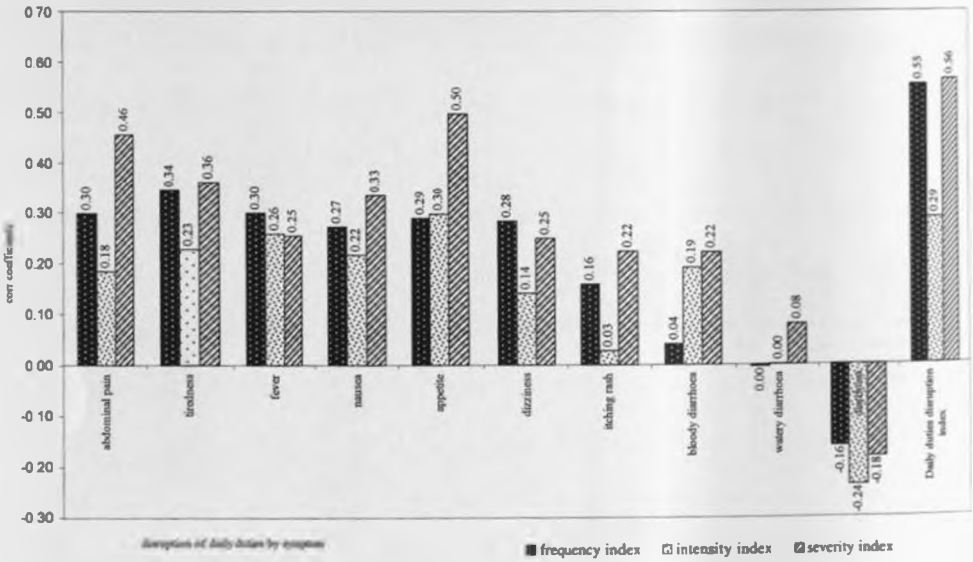


Figure A6.15: Correlation between infection intensity and how often HRQL domains were affected (patient group) [no correlations significant at p<0.05]

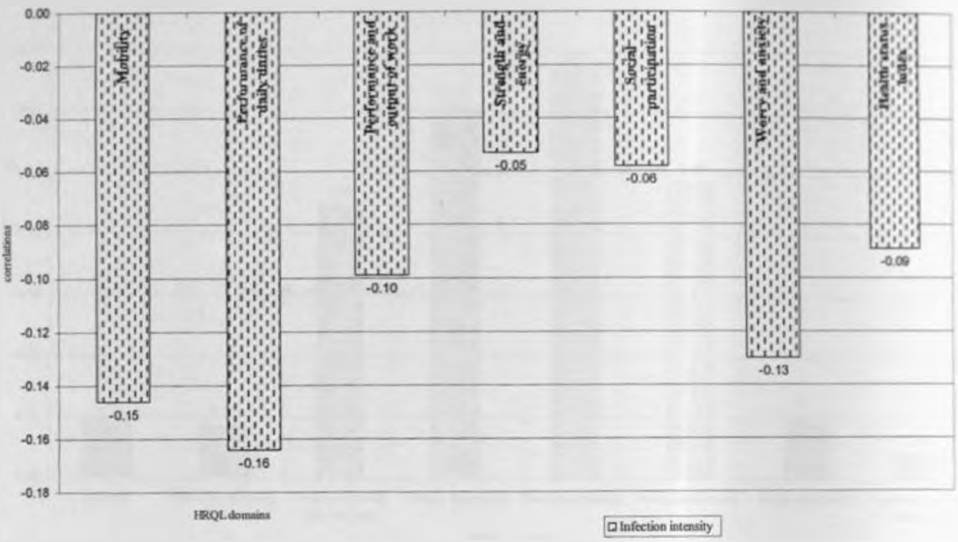


Figure A6.16: Correlation between infection intensity and how often daily duties were disrupted (patient group) [correlations >0.22 significant at p<0.05]

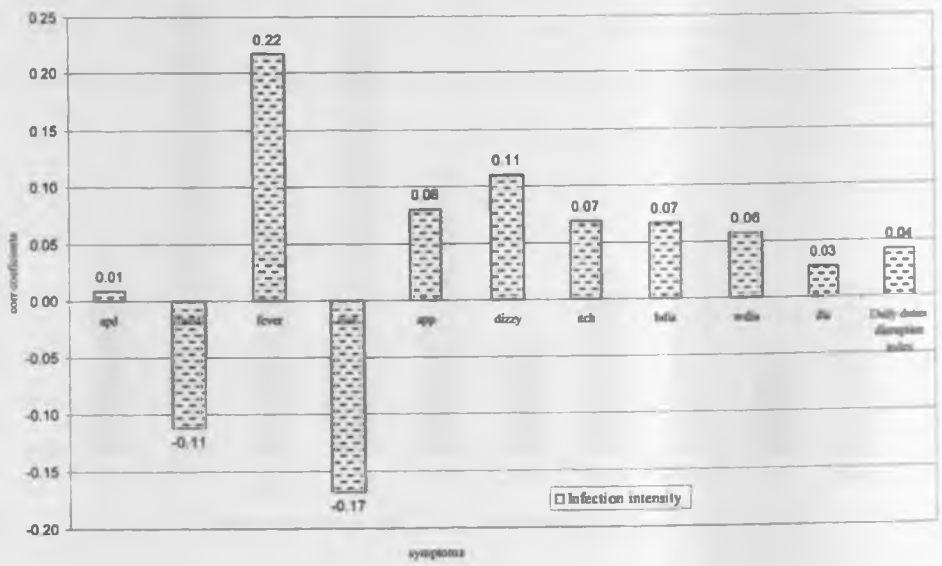


Figure A6.17: Kruskal Wallis test *p* values for association between infection intensity and HRQL indicators (patient group)

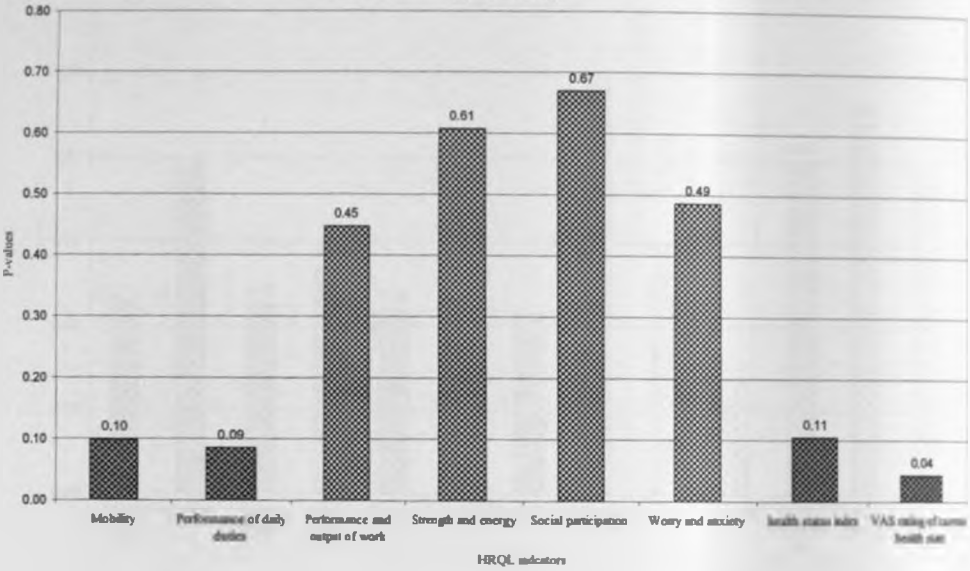


Figure A6.18: Kruskal Wallis test *p* values for association between infection intensity and disruption of daily duties (patient group)

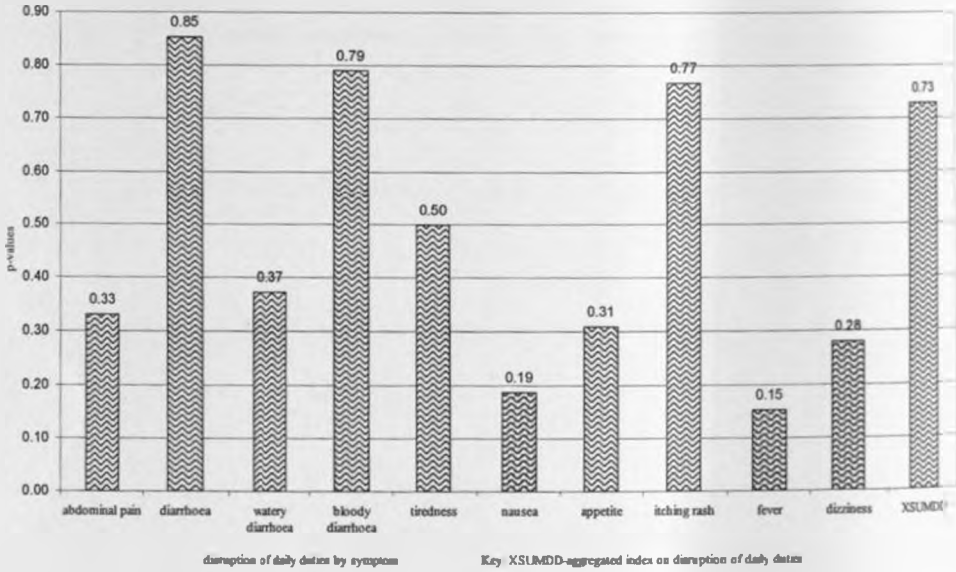
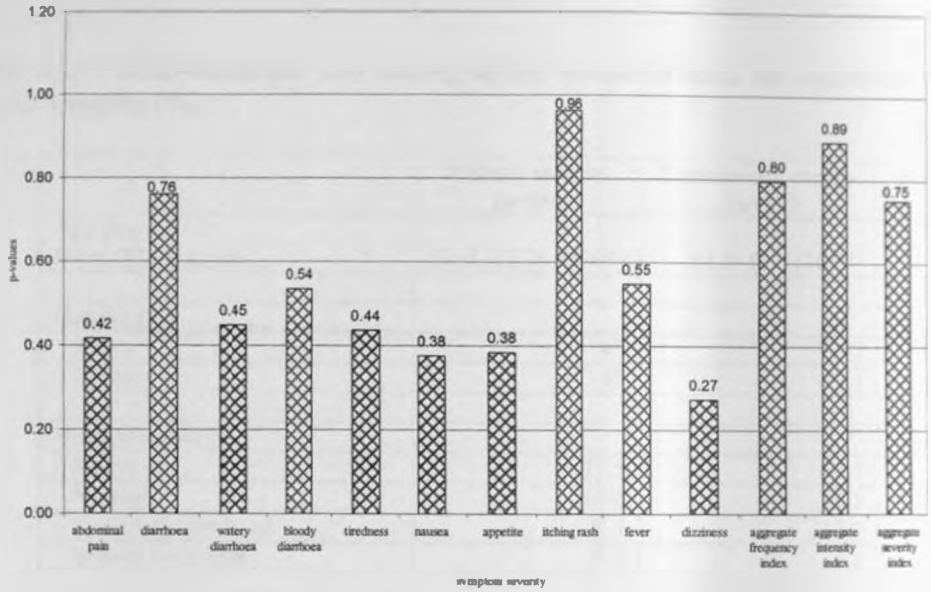


Figure A6.19: Kruskal Walis P values for infection intensity and severity of symptoms (patient group)



APPENDIX 6.2: TABLES

Table A6.1: Socio-economic and demographic characteristics for community and patient samples (%)

	Patient sample (n=80)	Community sample (n=81)	*P*-values
Age in years: Mean (SD) [range]	29.2 (12.7) [15-64]	41.2 (17) [16-77]	0.001
Gender (%)			
Males	45	37	
Females	55	63	0.338
Marital status (%)			
Single	37.5	11.1	0.001
Married	60.0	76.5	
Separated / divorced	2.5	3.7	
Widow / widower	0.0	8.6	
Education level (%)			
None	10	14.8	0.362
Primary	73.8	58.0	
Secondary (O and A level)	16.3	22.3	
Degree	0.0	1.2	
Adult education	0.0	3.7	
Occupation (%)			
Farmer	66.3	87.7	0.002
Teacher	0.0	2.5	
Business person	11.3	2.5	
Casual labourer	7.5	1.2	
Civil servant	1.3	0.0	
Student	13.8	2.5	
Other	0.0	2.5	

Table A6.2: Illness experienced in the last two weeks and during interview

	Patients (n=80)	Community (n=81)	*P* values
% reporting illness in the last 2 wks	67.5	60.5	0.413
% aware of which illness	42.5	59.3	0.378
% reporting illness during interview	92.5	32.1	0.000
% aware of which illness	76.3	29.6	0.000
% ever suffered from bilharzias before	86.3	54.3	0.000
% aware of illness with bilharzia	92.5	77.8	0.014
Illness experienced last 2 wks			
Bilharzia	6.3	4.9	
Malaria	10.0	14.8	
Typhoid	0.0	1.2	
Other worms	0.0	1.2	
Pains (different types)	5.0	16.0	
Bilharzia related symptoms	20.0	3.7	
Others	1.3	17.3	
Don't know	25.0	1.23	
Not ill	32.5	39.5	0.557
Illnesses experienced at the time of interview			
Bilharzia	70.0	1.2	
Malaria	0.0	6.2	
Typhoid	0.0	1.2	
Other worms	0.0	0.0	
Pains (different types)	0.0	8.6	
Bilharzia related symptoms	1.3	6.2	
Others	5.0	6.2	
Don't know	16.3	1.2	
Not ill	7.5	69.1	0.000

Table A6.3: Frequency, intensity and severity of symptoms (% by category)

Symptoms	frequency			Intensity			symptom seventy index (sum of frequency and intensity)		
	Patient	Community	difference in proportion (P-C)	Patient	Community	difference in proportion (P-C)	Patient	Community	difference in proportion (P-C)
Abdominal pain and discomfort									
1		2.5	-2.5	3.8	12.3	-8.5	2.5	9.9	-7.4
2	33.8	22.2	11.6	37.5	9.9	27.6	37.5	16	21.5
3	26.3	3.7	22.6	41.3	13.6	27.7	41.3	7.4	33.9
4	18.8	4.9	13.9	8.8	7.4	1.4	10	9.9	0.1
5	12.5	9.9	2.6						
***P* values**			0.000			0.000			0.000
Tiredness									
1		1.2	-1.2	5	4.9	0.1	2.5	4.9	-2.4
2	27.5	24.7	2.8	33.8	18.5	15.3	40	18.5	21.5
3	25	4.9	20.1	45	11.1	33.9	31.3	12.3	19
4	15	1.2	13.8	7.5	12.3	-4.8	17.5	11.1	6.4
5	23.8	14.8	9						
***P* values**			0.000			0.000			0.000
Fever									
1				2.5	2.5	0	1.3	1.2	0.1
2	33.8	11.1	22.7	25	4.9	20.1	37.5	9.9	27.6
3	30	9.9	20.1	50	19.8	30.2	36.3	12.3	24
4	8.8	2.5	6.3	10	4.9	5.1	12.5	8.6	3.9
5	15	8.6	6.4						
***P* values**			0.000			0.000			0.000
Nausea									
1		1.2	-1.2	8.8	11.1	-2.3	5	6.2	-1.2
2	28.8	13.6	15.2	31.3	2.5	28.8	40	9.9	30.1
3	31.3	7.4	23.9	35	14.8	20.2	23.8	13.6	10.2
4	8.8	4.9	3.9	5	3.7	1.3	11.3	2.5	8.8
5	11.3	4.9	6.4						
***P* values**			0.000			0.000			0.000
Loss of appetite									
1		1.2	-1.2	8.8		8.8	5	1.2	3.8
2	37.5	9.9	27.6	40	14.8	25.2	46.3	11.1	35.2
3	30	4.9	25.1	26.3	8.6	17.7	25	9.9	15.1
4	6.3	2.5	3.8	5	1.2	3.8	3.8	2.5	1.3
5	6.3	6.2	0.1						
***P* values**			0.000			0.000			0.000
Dizziness									
1		3.7	-3.7	5	6.2	-1.2	5	2.5	2.5
2	28.8	3.7	25.1	26.3	1.2	25.1	30	6.2	23.8
3	21.3	3.7	17.6	32.5	9.9	22.6	26.3	7.4	18.9
4	5	2.5	2.5	2.5	6.2	-3.7	5	7.4	-2.4
5	11.3	9.9	1.4						
***P* values**			0.000			0.000			0.000
Itching skin rash									
1				6.3	2.5	3.8	5		5
2	16.3	2.5	13.8	11.3	7.4	3.9	13.8	3.7	10.1
3	10	1.2	8.8	20	3.7	16.3	15	7.4	7.6
4	3.8	1.2	2.6	1.3	1.2	0.1	5	3.7	1.3
5	8.8	9.9	-1.1						
***P* values**			0.003			0.001			0.002

Bloody mucoid diarrhoea									
1		1.2	-1.2	26.3	6.2	20.1	16.3	4.9	11.4
2	21.3	3.7	17.6	3.8	1.2	2.6	15	2.5	12.5
3	10	2.5	7.5	6.3		6.3	5		5
4	3.8		3.8	1.3		1.3	1.3		1.3
5	2.5		2.5						
***P* values**			0.000			0.000			0.000
Watery diarrhoea									
1	1.3		1.3	20	1.2	18.8	10	1.2	8.8
2	8.8	1.2	7.6	2.5		2.5	10		16
3	10		10	2.5		2.5	5		5
4	3.8		3.8						
5	1.3		1.3						
***P* values**			0.000			0.000			0.000
Diarrhoea									
1				21.3	6.2	15.1	16.3	4.9	11.4
2	17.5	4.9	12.6	1.3		1.3	6.3	1.2	5.1
3	5	1.2	3.8						
4									
5									
***P* values**			0.001			0.003			0.003

1-5 frequency levels for frequency of symptoms (1= less than a day; 2=1-3 days; 3=4-6 days; 4=7-10 days; 5=11-14 days) 1-4 levels for intensity and severity of symptoms column (1=mild; 2=moderate; 3= severe; 4=very severe).
NB: figure reported are for the proportion reporting presence of symptom and therefore add up to % shown in table 5. *P* values are provided for Mann-Whitney test for differences between proportions of patients and community members on different levels of frequency, intensity and severity of symptoms.

Table A6.4: Whether symptom disrupts daily duties (%)[community, n=81; patients, n=80]

symptoms	Patients	Community	difference in proportions (P -C)
Abdominal pain and discomfort	80.0	27.2	52.8**
Tiredness	75.0	35.9	39.1**
Fever	62.5	27.2	35.3**
Nausea	62.5	17.3	45.2**
Loss of appetite	68.8	17.3	51.5**
Dizziness	57.5	17.3	40.2**
Itching skin rash	20.0	2.5	17.5**
Bloody mucoid diarrhoea	33.8	4.9	28.9**
Watery diarrhoea	21.3	1.2	20.1**
Diarrhoea	17.5	2.5	15.0**

Test statistic, Chi-square Fisher's exact test. *P values* ** <0.01

Table A6.5: How HRQL domains are affected by bilharzia (%)

Domain	How domain is affected	Patients (n=80)	Community (n=81)
Mobility	restricted in movement due to symptoms	57.5	43.2
	unable to walk due to body weakness	18.8	21.0
	when chronic one cannot move at all	2.5	
	mobility depends on severity of symptoms	11.3	
	Others	-	11.1
Performance of daily duties	inability to work due to symptoms	20.0	39.5
	inability to work due to body weakness	18.8	24.7
	inability to work due to illness	33.8	
	lack of progress in your work due to inability to go to work	8.8	9.9
	reduction in working hours	5.0	4.9
	others	7.5	4.9
Performance and output of one's work	reduction in performance and output due to illness	42.5	
	inability to work due to illness	25.0	32.1
	reduction in output due to body weakness	15.0	16.0
	misallocation of one's resources to cater for drugs and hospitalization	2.5	
	reduction in working hours due to absenteeism	2.5	23.5
	others	2.5	6.2
Feeling of strength and energy in body	inability to work due to body weakness	10.0	9.9
	one's effort to work is brought to a minimal	7.5	
	restricts mobility due to body weakness	1.3	
	bilharzia drinks your blood and eats your body	2.5	4.9
	leads to body weakness	48.8	53.1
	bilharzia kills body joints and body cells	15.0	
	others	10.0	13.6
Social participation	restrictions in socialisation due to discomfort from symptoms	46.3	42.0
	lack of concentration when socialising due to symptoms	3.8	3.7
	lack of strength to go to social places	10.0	16.0
	feelings of inadequacy due to body appearance	5.0	
	lack of happiness and joy to socialise	8.8	9.9
	others	6.3	1.2
Feelings of worry and anxiety	worry because of illness and symptoms	25.0	25.9
	worry about your life being in dangers because of the illness	5.0	
	worry about where to get resources to secure treatment	7.5	7.4
	worry because you are not doing your work	15.0	12.3

	worry about whether you will get cured	5.0	7.4
	worry due to lack of happiness and body strength	7.5	
	worry about the likelihood of dying due to illness	2.5	8.6
	bilharzia affects other family members	1.3	
	suspicious about type and cause of illness	15.0	7.4
	worry due to restlessness of mind	3.8	
	Others		

Table A6.6: Summary of evidence of concurrent validity: Significant correlations between HRQL indicators and symptoms ($p < 0.05$)

Associations tested	Abdominal pain and discomfort	Tiredness	Fever	Nausea	Loss of appetite	Dizziness	Itching skin rash	Bloody diarrhoea	Watery diarrhoea	Diarrhoea	XFRQ	XINTE	XFRIN
Association between frequency of symptoms and how often HRQL domains were affected by schistosomiasis mansoni		Perf Ener soc		wory	Perf soc	Soc							
Association between intensity of symptoms and how often HRQL domains were affected by schistosomiasis mansoni		Mob Perf ener					Perf						
Association between severity of symptoms and how often HRQL domains were affected by schistosomiasis mansoni		Mob Perf Ener Soc			Perf Ener Soc	Soc							
Association between aggregate frequency, intensity and severity of symptoms and how often HRQL domains were affected by schistosomiasis mansoni											Perf Soc wory		Soc
Association between aggregate frequency, intensity and severity of symptoms and health status index											Hesi		

Association between frequency of symptoms and ways in which HRQL domains were affected by schistosomiasis mansoni		wory	Mob Perf ener	Ener						wory			
Association between intensity of symptoms and ways in which HRQL domains were affected by schistosomiasis mansoni	perf	wory	Mob Perf	perf				perf	Mob	Work			
Association between severity of symptoms and ways in which HRQL domains were affected by schistosomiasis mansoni		Mob Soc wory	Ener		Soc			Wory					
Association between aggregate frequency, intensity and severity of symptoms and ways in which HRQL domains were affected by schistosomiasis mansoni												Mob Perf wory	Perf Work

Key: Mob-mobility. Perf-performance of daily duties. Work-performance and output of work. Ener-strength and energy. Soc-social participation. wory-worry and anxiety. Hesi-health status index. XFRQ- overall frequency index . XINTE-overall intensity index. XFRIN-overall severity index.

Table A6.7: Kruskal Wallis test *p* values for tests of differences between categories of severity of symptoms and HRQL indicators

HRQL indicators	Symptom severity indices												
	Abdominal pain and discomfort	Diarrhoea	Watery diarrhoea	Bloody diarrhoea	Tiredness	Nausea	Loss of appetite	Itching skin rash	Fever	Dizziness	Total frequency index	Total intensity index	Total severity index
Mobility	0.471	0.284	0.402	0.219	0.165	0.543	0.322	0.898	0.144	0.943	0.156	0.566	0.172
performance of daily duties	0.484	0.172	0.826	0.143	0.087	0.334	**0.031**	0.472	0.702	0.833	**0.029**	0.076	**0.027**
performance and output of work	0.246	0.69	0.677	0.381	0.793	0.183	0.124	0.914	0.857	0.131	0.404	0.39	0.324
energy and strength	0.176	0.094	0.737	0.36	0.069	0.347	0.37	0.427	0.278	0.436	0.184	0.152	0.201
social participation	0.533	0.654	0.665	**0.043**	0.1	0.348	**0.012**	0.529	0.369	0.097	**0.036**	0.136	**0.029**
worry and anxiety	0.278	0.458	0.275	**0.009**	0.503	0.527	0.346	0.314	0.123	0.491	**0.003**	0.586	0.188
Health status index	**0.05**	0.11	0.615	0.444	0.195	0.262	**0.027**	0.668	0.388	0.157	0.074	0.465	0.18
VAS rating of current health state	0.072	0.158	**0.035**	**0.002**	0.072	0.149	0.067	0.434	0.556	0.669	0.164	0.219	**0.012**

In bold: $p \leq 0.05$

Table A6.8: Cramer's V correlation coefficients between frequency, intensity and severity of symptoms and ways HRQL domains are affected by bilharzia [significance level]

Symptoms		How bilharzia affects HRQL domains					
		Mobility	Performance of daily duties	Performance and output of work	Strength and energy	Social participation	Worry and anxiety
Abdominal pain and discomfort	Frequency	0.223 [0.459]	0.300 [0.425]	0.321 [0.238]	0.312 [0.512]	0.265 [0.758]	0.422 [0.089]
	Intensity	.0233 [0.360]	**0.383 [0.014]**	0.314 [0.293]	0.368 [0.088]	0.277 [0.654]	0.370 [0.483]
	Severity index	0.209 [0.599]	0.266 [0.754]	0.288 [0.546]	0.310 [0.529]	0.262 [0.783]	0.396 0.240]
Diarrhoea	Frequency	0.188 [0.689]	0.263 [0.681]	0.303 [0.401]	0.258 [0.831]	0.239 [0.821]	**0.507 [0.008]**
	Intensity	0.186 [0.696]	0.230 [0.862]	**0.518 [0.000]**	0.250 [0.868]	0.225 [0.868]	0.388 [0.341]
	Severity index	0.186 [0.700]	0.246 [0.786]	0.335 [0.207]	0.274 [0.745]	0.222 [0.894]	0.472 [0..34]
Watery diarrhoea	Frequency	0.295 [0.115]	0.299 [0.437]	0.263 [0.809]	0.308 [0.563]	0.308 [0.335]	0.270 [0.998]
	Intensity	**0.316 [0.020]**	0.245 [0.853]	0.194 [0.989]	0.241 [0.948]	0.246 [0.848]	0.281 [0.976]
	Severity index	0.249 [0.249]	0.229 [0.923]	0.222 [0.943]	0.207 [0.993]	0.328 [0.214]	0.287 [0.966]
Bloody diarrhoea	Frequency	0.173 [0.888]	0.296 [0.467]	0.265 [0.761]	0.296 [0.667]	0.236 [0.931]	0.390 [0.294]
	Intensity	0.238 [0.314]	**0.397 [0.006]**	0.244 [0.898]	0.240 [0.973]	0.278 [0.645]	0.420 [0.100]
	Severity index	0.228 [0.413]	0.336 [0.141]	0.234 [0.938]	0.282 [0.788]	0.267 [0.739]	**0.440 [0.038]**
Tiredness	Frequency	0.191 [0.765]	0.287 [0.558]	0.303 [0.394]	**0.386 [0.037]**	0.305 [0.378]	**0.467 [0.008]**
	Intensity	0.229 [0.395]	0.258 [0.814]	0.284 [0.558]	0.355 [0.147]	0.306 [0.363]	**0.497 [0.001]**
	Severity index	**0.290 [0.042]**	0.278 [0.644]	0.332 [0.162]	0.355 [0.148]	**0.363 [0.041]**	**0.474 [0.005]**
Nausea	Frequency	0.235 [0.341]	0.266 [0.754]	0.312 [0.306]	**0.403 [0.015]**	0.276 [0.665]	0.385 [0.337]
	Intensity	0.267 [0.121]	**0.372 [0.026]**	0.319 [0.250]	0.293 [0.697]	0.239 [0.919]	0.336 [0.798]
	Severity index	0.223 [0.455]	0.311 [0.318]	0.307 [0.358]	0.338 [0.265]	0.254 [0.843]	0.374 [0.444]
Loss of appetite	Frequency	0.189 [0.784]	0.237 [0.925]	0.252 [0.852]	0.281 [0.791]	0.311 [0.315]	0.420 [0.097]
	Intensity	0.257 [0.175]	0.292 [0.501]	0.318 [0.256]	0.315 [0.483]	0.279 [0.638]	0.423 [0.088]
	Severity index	0.247 [0.240]	0.346 [0.094]	0.324 [0.215]	0.295 [0.674]	**0.366 [0.036]**	0.394 [0.261]

Itching skin rash	Frequency	0.209 [0.598]	0.322 [0.227]	0.250 [0.863]	0.296 [0.667]	0.294 [0.480]	0.371 [0.467]
	Intensity	0.236 [0.336]	0.339 [0.126]	0.220 [0.972]	0.327 [0.362]	0.281 [0.616]	0.321 [0.887]
	Severity index	0.210 [0.589]	0.326 [0.199]	0.301 [0.411]	0.309 [0.541]	0.313 [0.306]	0.362 [0.558]
Fever	Frequency	**0.287 [0.049]**	**0.361 [0.046]**	0.263 [0.773]	**0.388 [0.033]**	0.297 [0.456]	0.353 [0.649]
	Intensity	**0.294 [0.035]**	**0.413 [0.002]**	0.315 [0.284]	0.353 [0.162]	0.304 [0.384]	0.412 [0.139]
	Severity index	0.250 [0.180]	0.351 [0.075]	0.245 [0.891]	**0.403 [0.014]**	0.259 [0.803]	0.386 [0.323]
Dizziness	Frequency	0.174 [0.884]	0.268 [0.731]	0.301 [0.418]	0.322 [0.407]	0.256 [0.823]	0.410 [0.145]
	Intensity	0.207 [0.619]	0.297 [0.455]	0.280 [0.624]	0.297 [0.654]	0.293 [0.493]	0.367 [0.507]
	Severity index	0.163 [0.934]	0.216 [0.980]	0.271 [0.705]	0.301 [0.621]	0.265 [0.759]	0.420 [0.098]

Table A6.9: Spearman's rho rank correlation coefficients between frequency, intensity and severity of symptoms and how often symptom disrupts daily duties [significance level]

Symptoms		How often daily duties are disrupted by symptoms										Daily duties disruption index
		Abdominal pain and discomfort	Diarrhoea	Watery diarrhoea	Bloody diarrhoea	Tiredness	Nausea	Loss of appetite	Itching skin rash	Fever	Dizziness	
Abdominal pain and discomfort	Frequency	**0.777 [0.000]**	0.029 [0.796]	0.183 [0.105]	**0.295 [0.008]**	0.139 [0.221]	0.164 [0.146]	**0.338 [0.002]**	0.054 [0.632]	0.008 [0.944]	0.046 [0.684]	**0.231 [0.040]**
	Intensity	**0.461 [0.000]**	0.083 [0.462]	0.128 [0.259]	**0.324 [0.003]**	0.085 [0.454]	0.109 [0.336]	**0.239 [0.033]**	0.183 [0.105]	-0.061 [0.591]	0.072 [0.525]	0.087 [0.441]
	Severity index	**0.741 [0.000]**	0.013 [0.906]	0.196 [0.081]	**0.340 [0.002]**	0.131 [0.246]	0.131 [0.247]	**0.383 [0.000]**	0.070 [0.535]	-0.043 [0.705]	0.053 [0.643]	0.209 [0.062]
Diarrhoea	Frequency	-0.083 [0.466]	**0.868 [0.000]**	-0.186 [0.099]	**-0.264 [0.018]**	-0.123 [0.279]	0.161 [0.154]	**-0.219 [0.052]**	-0.045 [0.692]	0.036 [0.745]	0.076 [0.504]	**-0.231 [0.039]**
	Intensity	-0.081 [0.473]	**0.853 [0.000]**	-0.201 [0.074]	**-0.273 [0.014]**	-0.118 [0.295]	0.159 [0.158]	**-0.234 [0.037]**	-0.052 [0.648]	0.031 [0.787]	0.067 [0.552]	**-0.228 [0.042]**
	Severity index	-0.072 [0.523]	**0.870 [0.000]**	-0.187 [0.097]	**-0.265 [0.017]**	-0.124 [0.273]	0.170 [0.131]	-0.215 [0.055]	0.048 [0.674]	0.045 [0.690]	0.077 [0.499]	**-0.219 [0.051]**
Watery diarrhoea	Frequency	0.202 [0.073]	-0.173 [0.124]	**0.918 [0.000]**	-0.025 [0.826]	0.075 [0.509]	-0.148 [0.190]	**0.249 [0.026]**	0.009 [0.936]	-0.019 [0.865]	-0.094 [0.409]	0.065 [0.568]
	Intensity	0.180 [0.111]	-0.183 [0.105]	**0.905 [0.000]**	-0.019 [0.867]	0.062 [0.585]	-0.164 [0.145]	**0.223 [0.047]**	0.023 [0.838]	-0.014 [0.902]	-0.106 [0.348]	0.046 [0.682]
	Severity index	0.201 [0.074]	-0.173 [0.124]	**0.913 [0.000]**	-0.019 [0.868]	0.069 [0.543]	-0.151 [0.180]	**0.245 [0.029]**	0.009 [0.937]	-0.022 [0.847]	-0.096 [0.396]	0.056 [0.623]
Bloody diarrhoea	Frequency	**0.270 [0.015]**	-0.146 [0.197]	-0.012 [0.916]	**0.921 [0.000]**	-0.011 [0.925]	0.144 [0.201]	**0.222 [0.048]**	0.122 [0.281]	0.139 [0.220]	0.003 [0.979]	0.124 [0.274]
	Intensity	**0.232 [0.038]**	-0.159 [0.158]	-0.028 [0.805]	**0.890 [0.000]**	0.004 [0.973]	0.120 [0.289]	0.198 [0.079]	0.137 [0.225]	0.105 [0.354]	-0.022 [0.846]	0.092 [0.415]
	Severity index	**0.263 [0.018]**	-0.156 [0.167]	-0.018 [0.875]	**0.907 [0.000]**	-0.007 [0.950]	0.128 [0.259]	**0.215 [0.055]**	0.116 [0.306]	0.134 [0.236]	-0.010 [0.931]	0.117 [0.301]
Tiredness	Frequency	0.214 [0.056]	-0.100 [0.374]	-0.061 [0.592]	0.119 [0.294]	**0.619 [0.000]**	0.209 [0.063]	**0.253 [0.024]**	0.118 [0.299]	**0.218 [0.052]**	0.202 [0.073]	**0.397 [0.000]**
	Intensity	**0.238 [0.034]**	0.044 [0.698]	0.114 [0.313]	0.113 [0.320]	**0.489 [0.000]**	**0.286 [0.010]**	**0.278 [0.013]**	-0.026 [0.817]	**0.303 [0.006]**	**0.233 [0.037]**	**0.347 [0.002]**
	Severity index	**0.255 [0.022]**	-0.102 [0.368]	-0.015 [0.898]	0.161 [0.153]	**0.613 [0.000]**	**0.243 [0.030]**	**0.280 [0.012]**	0.091 [0.424]	**0.259 [0.020]**	**0.225 [0.045]**	**0.444 [0.000]**

Nausea	Frequency	**0.247 [0.027]**	0.145 [0.200]	-0.160 [0.157]	0.165 [0.145]	0.104 [0.361]	**0.663 [0.000]**	**0.266 [0.017]**	-0.003 [0.980]	0.184 [0.103]	**0.367 [0.001]**	**0.238 [0.034]**
	Intensity	0.173 [0.125]	**0.245 [0.029]**	-0.136 [0.230]	0.190 [0.091]	0.033 [0.771]	**0.596 [0.000]**	**0.338 [0.002]**	0.009 [0.937]	**0.267 [0.017]**	**0.346 [0.002]**	0.139 [0.219]
	Severity index	**0.231 [0.040]**	0.158 [0.161]	-0.154 [0.172]	0.189 [0.093]	0.051 [0.653]	**0.689 [0.000]**	**0.338 [0.002]**	-0.047 [0.680]	**0.252 [0.024]**	**0.362 [0.001]**	**0.223 [0.047]**
Loss of appetite	Frequency	**0.342 [0.002]**	-0.124 [0.273]	0.179 [0.113]	**0.273 [0.014]**	**0.266 [0.017]**	**0.431 [0.000]**	**0.762 [0.000]**	0.212 [0.059]	**0.217 [0.053]**	**0.315 [0.004]**	**0.385 [0.000]**
	Intensity	**0.287 [0.010]**	-0.079 [0.488]	0.201 [0.074]	**0.253 [0.024]**	0.139 [0.218]	**0.285 [0.010]**	**0.608 [0.000]**	0.105 [0.354]	0.111 [0.326]	0.143 [0.205]	**0.217 [0.054]**
	Severity index	**0.357 [0.001]**	-0.070 [0.534]	0.183 [0.104]	**0.278 [0.012]**	**0.254 [0.023]**	**0.376 [0.001]**	**0.756 [0.000]**	0.196 [0.082]	0.133 [0.238]	**0.295 [0.008]**	**0.352 [0.001]**
Itching skin rash	Frequency	-0.114 [0.315]	0.205 [0.069]	0.063 [0.576]	0.123 [0.276]	0.046 [0.685]	0.069 [0.541]	0.063 [0.578]	**0.710 [0.000]**	-0.113 [0.318]	-0.003 [0.981]	-0.032 [0.779]
	Intensity	-0.137 [0.227]	0.208 [0.064]	0.077 [0.496]	0.098 [0.385]	0.050 [0.660]	0.082 [0.472]	0.077 [0.498]	**0.647 [0.000]**	-0.043 [0.704]	-0.034 [0.767]	-0.023 [0.837]
	Severity index	-0.124 [0.273]	**0.217 [0.053]**	0.044 [0.698]	0.113 [0.319]	0.052 [0.645]	0.085 [0.451]	0.071 [0.534]	**0.696 [0.000]**	-0.087 [0.444]	-0.007 [0.952]	-0.026 [0.821]
Fever	Frequency	0.041 [0.719]	0.182 [0.106]	-0.082 [0.470]	0.167 [0.140]	0.156 [0.166]	**0.354 [0.001]**	0.121 [0.286]	0.040 [0.722]	**0.653 [0.000]**	**0.322 [0.004]**	**0.354 [0.001]**
	Intensity	0.059 [0.603]	-0.089 [0.432]	-0.058 [0.608]	**0.257 [0.021]**	**0.250 [0.026]**	0.138 [0.222]	0.163 [0.149]	0.094 [0.405]	**0.559 [0.000]**	0.192 [0.087]	**0.287 [0.010]**
	Severity index	0.102 [0.366]	0.096 [0.396]	-0.089 [0.430]	0.174 [0.122]	**0.238 [0.033]**	**0.359 [0.001]**	0.166 [0.142]	0.113 [0.319]	**0.717 [0.000]**	**0.328 [0.003]**	**0.429 [0.000]**
Dizziness	Frequency	0.137 [0.224]	0.085 [0.451]	-0.076 [0.503]	0.049 [0.668]	**0.236 [0.035]**	**0.369 [0.001]**	**0.289 [0.009]**	0.055 [0.628]	**0.234 [0.036]**	**0.822 [0.000]**	**0.290 [0.009]**
	Intensity	0.149 [0.186]	0.114 [0.316]	-0.024 [0.831]	0.036 [0.752]	0.212 [0.059]	**0.374 [0.001]**	**0.273 [0.014]**	-0.053 [0.641]	**0.306 [0.006]**	**0.747 [0.000]**	**0.233 [0.038]**
	Severity index	0.183 [0.105]	0.067 [0.557]	-0.061 [0.589]	0.062 [0.587]	**0.248 [0.027]**	**0.385 [0.000]**	**0.305 [0.006]**	0.019 [0.869]	**0.302 [0.006]**	**0.817 [0.000]**	**0.305 [0.006]**

Table A6.10: Kruskal Wallis test *p* values for tests of differences between categories of severity of symptoms and disruption of daily duties

Disruption of daily duties	Symptom severity indices												
	Abdominal pain and discomfort	Diarrhoea	Watery diarrhoea	Bloody diarrhoea	Tiredness	Nausea	Loss of appetite	Itching skin rash	Fever	Dizziness	Total frequency index	Total intensity index	Total severity index
Abdominal pain and discomfort	**0.000**	0.248	**0.009**	0.079	0.137	0.058	**0.036**	0.725	0.591	0.218	**0.005**	0.263	**0.000**
Diarrhoea	0.512	**0.000**	0.362	0.705	0.594	0.115	0.928	**0.046**	0.916	**0.02**	0.528	0.108	0.455
Watery diarrhoea	0.45	0.127	**0.000**	0.979	0.814	0.54	0.097	0.398	0.636	0.636	0.905	0.257	0.712
Bloody diarrhoea	**0.028**	**0.049**	0.694	**0.000**	0.61	0.332	0.136	0.34	0.441	0.817	0.771	0.1	0.138
Tiredness	**0.049**	0.488	0.153	0.661	**0.000**	0.154	0.109	0.22	0.177	0.077	**0.026**	**0.008**	**0.001**
Nausea	0.244	0.108	0.347	0.439	0.252	**0.000**	**0.002**	0.547	**0.021**	**0.006**	0.168	**0.029**	**0.005**
Loss of appetite	**0.013**	**0.032**	**0.012**	0.078	**0.045**	**0.031**	**0.000**	0.932	0.392	0.061	**0.004**	**0.004**	**0.000**
Itching skin rash	0.534	0.905	0.236	0.403	0.65	0.255	0.301	**0.000**	0.258	0.279	0.504	0.679	0.138
Fever	0.935	0.175	0.813	0.263	0.193	0.254	0.784	0.433	**0.000**	**0.039**	0.099	**0.006**	**0.028**
Dizziness	0.981	0.596	0.789	0.419	**0.012**	**0.032**	**0.033**	0.684	**0.062**	**0.000**	0.073	0.445	**0.042**
Total disruption of daily duties index	0.299	**0.033**	0.117	0.178	**0.003**	0.299	**0.019**	0.489	**0.000**	**0.023**	**0.000**	**0.004**	**0.000**

In bold: $p \leq 0.05$

APPENDIX 6.3: SPECIFIC HYPOTHESIS FOR ASSESSING CONSTRUCT VALIDITY

1. Patients with higher infection intensity have worse off HRQL indicators. This hypothesis was tested by running the following sub hypothesis:
    - Patients with higher infection intensity have worse off health status index
    - Patients with higher infection intensity report that bilharzia affects their HRQL domains more often.
    - Patients with higher infection intensity report more that their daily duties are disrupted more often.
    - Patients with higher infection intensity report lower VAS rating of current health status.

2. Patients with higher infection intensity have higher symptom severity index.
This hypothesis was tested using the following sub hypotheses:

    - Patients with higher infection intensity have a higher individual symptom severity index.
    - Patients with higher infection intensity have a higher aggregate frequency index i.e. report more frequency of symptoms.
    - Patients with higher infection intensity have a higher aggregate intensity index i.e. report more intensity on symptoms.
    - Patients with higher infection intensity have a higher aggregate symptom severity index.

3. The more severe the symptoms the more HRQL domains are affected.

This hypothesis was tested using the following sub hypothesis:

- The higher the individual symptoms severity the worse off the health status index.
- The higher the aggregate symptom frequency the worse off the health status index
- The higher the aggregate symptom intensity index the worse off the health status index
- The higher the aggregate symptom severity index the worse off the health status index.

- The higher the individual symptoms severity the higher the reported frequency of how bilharzia affects HRQL domains
- The higher the aggregate symptom frequency the higher the reported frequency of how bilharzia affects HRQL domains
- The higher the aggregate symptom intensity index the higher the reported frequency of how bilharzia affects HRQL domains
- The higher the aggregate symptom severity index the higher the reported frequency of how bilharzia affects HRQL domains.

- The higher the individual symptoms severity the higher the reported frequency of disruption of daily duties.
- The higher the aggregate symptom frequency higher the reported frequency of disruption of daily duties.
- The higher the aggregate symptom intensity index the higher the reported frequency of disruption of daily duties.
- The higher the aggregate symptom severity index the higher the reported frequency of disruption of daily duties.

- The higher the individual symptoms severity the higher the aggregate index of frequency of disruption of daily duties.
- The higher the aggregate symptom frequency higher the aggregate index of frequency of disruption of daily duties.
- The higher the aggregate symptom intensity index the higher the aggregate index of frequency of disruption of daily duties.
- The higher the aggregate symptom severity index the higher the aggregate index of frequency of disruption of daily duties.

- The higher the individual symptoms severity the lower the VAS rating of current health state.
- The higher the aggregate symptom frequency higher the lower the VAS rating of current health state.
- The higher the aggregate symptom intensity y index the lower the VAS rating of current health state.
- The higher the aggregate symptom severity index the lower the VAS rating of current health state.

APPENDIX 7.1: CONSTRUCTION OF DISEASE STATES AFTER FINAL ANALYSIS OF PATIENT RESPONSES (HRQL DOMAINS AND SYMPTOMS AFFECTING THEM)

MOBILITY (Tiredness, fever, watery diarrhea)

1. You feel somewhat tired for 1-3 days; have moderate fever for 1-3 days and this affects your mobility a little of the time.

2. You feel very tired for 4-6 days; have severe fever for 4-6 days and this affects your mobility some of the time.

3. You feel very tired for 7-10 days; have severe fever for 4-6 days; and this affects your mobility some of the time.

PERFORMANCE OF DAILY DUTIES (Tiredness, loss of appetite, itching skin rash fever, bloody mucoid diarrhea, nausea and abdominal pain and discomfort)

1. You feel somewhat tired for 1-3 days; can eat ¾ of the food you normally eat for 1-3 days; have moderate fever for 1-3 days; feel mild nausea for 1-3 days and have moderate abdominal pain and discomfort for 1-3 days; and this affects your performance of daily duties a little of the time.

2. You feel very tired for 4-6 days; can eat ½ -¾ of the food you normally eat for 1-3 days; have moderate fever for 4-6 days; feel moderate nausea for 4-6 days and have moderate abdominal pain and discomfort for 4-6 days; and this affects your performance of daily duties some of the time.

3. You feel very tired for 7-10 days; can eat not more than two spoonfuls of the food you normally eat for 4-6 days; have moderate itching skin rash for 1-3 days; have blood mucoid diarrhea 3-9 time a day for 1-3 days; have moderate fever for 4-6 days; feel severe nausea for 4-6 days and have severe abdominal pain and discomfort for 7-10 days; and this affects your performance of daily duties some of the time.

FEELING OF ENERGY AND STRENGTH (Tiredness, loss of appetite and fever)

1. You feel somewhat tired for 1-3 days; can eat ¾ of the food you normally eat for 1-3 days; have moderate fever for 1-3 days; and this affects your feeling of strength and energy a little of the time.

2. You feel very tired for 4-6 days; can eat ½ - ¾ of the food you normally eat for 1-3 days; have moderate fever for 1-3 days; and this affects your feeling of strength and energy some of the time.

3. You feel very tired for 7-10 days; can no more than two spoonfuls of the food you normally eat for 4-6 days; have moderate fever for 1-3 days; and this affects your feeling of strength and energy some of the time.

SOCIAL PARTICIPATION (Tiredness, loss of appetite and Dizziness)

1. You feel somewhat tired for 1-3 days; can eat ¾ of the food you normally eat for 1-3 days; and this affects your social participation a little of the time.

2. You feel very tired for 4-6 days; can eat ½ - ¾ of the food you normally eat for 1-3 days; feel moderate dizziness for 1-3 days; and this affects your social participation some of the time.

3. You feel very tired for 7-10 days; can no more than two spoonfuls of the food you normally eat for 4-6 days; feel severe dizziness for 4-6 days; and this affects your social participation some of the time.

WORRY AND ANXIETY (Tiredness, bloody mucoid diarrhea and nausea)

1. You feel somewhat tired for 1-3 days; feel mild nausea for 1-3 days; and this causes feelings of worry and anxiety a little of the time.

2. You feel very tired for 4-6 days; feel moderate nausea for 4-6 days; and this causes feelings of worry and anxiety some of the time.

3. You feel very tired for 7-10 days; have blood mucoid diarrhea 3-9 time a day for 1-3 days; feel severe nausea for 4-6 days; and this causes feelings of worry and anxiety some of the time.

SIX SELECTED HEALTH STATES

STATE 11111	A
You feel somewhat tired for 1-3 days	
Can eat ¾ of the food you normally eat for 1-3 days	
Have moderate fever for 1-3 days	
Feel mild nausea for 1-3 days	
Have moderate abdominal pain and discomfort for 1-3 days	
Your mobility, performance of daily duties, feeling of strength and energy, social participation, feelings of worry and anxiety are affected a little of the time	

STATE 12232 **B**

You feel somewhat tired for 1-3 days, then very tired for 4-10 days

Can eat ½ -¾ of the food you normally eat for 1-3 days and then no more than two spoonfuls of the food you normally eat for 4-6 days

Have moderate fever for 1-6 days

Feel moderate nausea for 4-6 days

Have moderate abdominal pain and discomfort for 4-6 days

Feel severe dizziness for 4-6 days

Your mobility is affected a little of the time while performance of daily duties, feeling of strength and energy, social participation and feelings of worry and anxiety are affected some of the time.

STATE 21102 **C**

You feel somewhat tired for 1-3 days and then very tired for 4-6 days

Can eat ¾ of the food you normally eat for 1-3 days

Have moderate fever for 1-3 days and then severe fever for 4-6 days

Feel mild nausea for 1-3 days and then moderate nausea for 4-6 days

Have moderate abdominal pain and discomfort for 1-3 days

Your performance of daily duties and feeling of strength and energy are affected a little of the time while mobility and feelings of worry and anxiety are affected some of the time. Your social participation is not affected at all.

STATE 22222 **D**

You feel very tired for 4-6 days

Can eat ½ -¾ of the food you normally eat for 1-3 days

Have moderate to severe fever for 1-6 days

Feel moderate nausea for 4-6 days

Have moderate abdominal pain and discomfort for 4-6 days

Feel moderate dizziness for 1-3 days

Your mobility, performance of daily duties, feeling of strength and energy, social participation and feelings of worry and anxiety are affected some of the time.

STATE 23222 **E**

You feel very tired for 4-10 days

Can eat ½ - ¾ of the food you normally eat for 1-3 days and then no more than two spoonfuls of the food you normally eat for 4-6 days

Have moderate to severe fever for 1-6 days

Have moderate itching skin rash for 1-3 days

Have blood mucoid diarrhea 3-9 time a day for 1-3 days

Feel moderate to severe nausea for 4-6 days

Have severe abdominal pain and discomfort for 7-10 days

Feel moderate dizziness for 1-3 days

Your mobility, performance of daily duties, feeling of strength and energy, social participation and feelings of worry and anxiety are affected some of the time.

STATE 33333 **F**

You feel very tired for 7-10 days

Can eat not more than two spoonfuls of the food you normally eat for 4-6 days

Have moderate to severe fever for 1-6 days

Have moderate itching skin rash for 1-3 days

Have blood mucoid diarrhea 3-9 time a day for 1-3 days

Feel severe nausea for 4-6 days

Have severe abdominal pain and discomfort for 7-10 days

Feel severe dizziness for 4-6 days

Your mobility, performance of daily duties, feeling of strength and energy, social participation and feelings of worry and anxiety are affected some of the time.

APPENDIX 7.2: HRQL DOMAIN AND SYMPTOMS: PERCENTILES FOR CONSTRUCTING DISEASE STATES

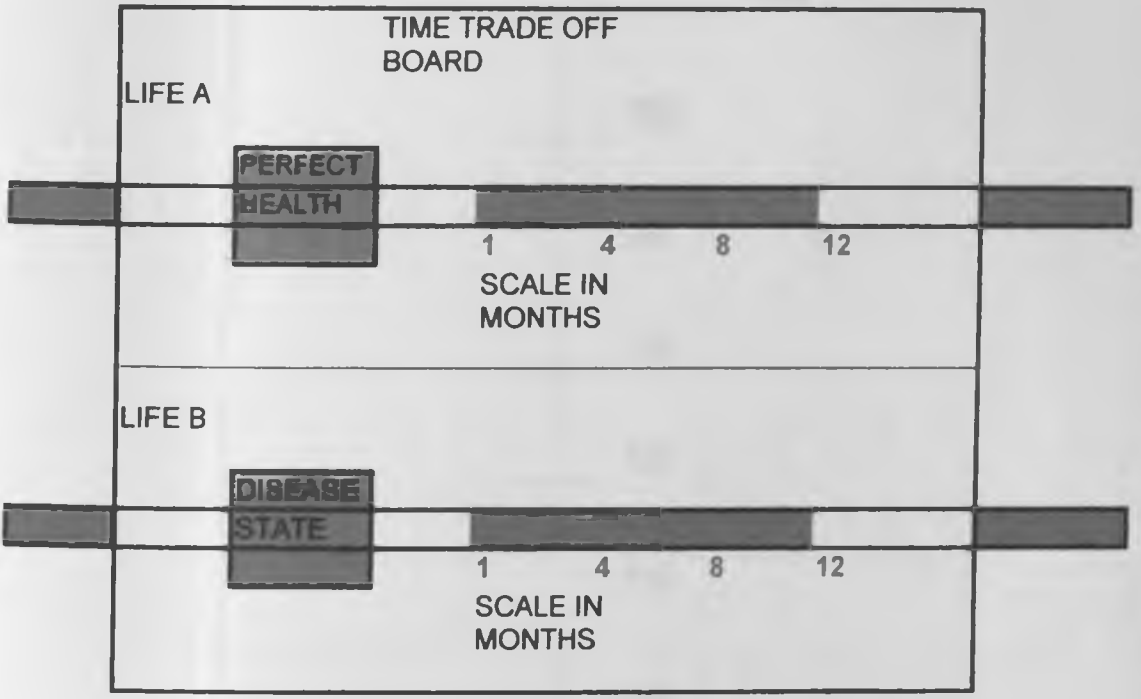
		How often mobility affected	How often performance of daily duties is affected	How often performance and output of one's work is affected	How often feeling of energy and strength in body is affected	How often ability to socialize is affected	How often bilharzia causes worry and anxiety	Number of days you had apd in the last two weeks	Watery diarrhoea in the last two weeks	Bloody diarrhoea in the last two weeks	Tiredness in the last two weeks	Nausea in the last two weeks	Loss of appetite in the last two weeks	Itching skin rash in the last two weeks	Fever in the last two weeks	Dizziness in the last two weeks
N	Valid	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
	Missing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Median		2.00	2.00	2.00	2.00	2.00	2.00	3.00	.00	.00	3.00	3.00	2.00	.00	3.00	2.00
Mode		2.00	2.00	2.00	2.00	2.00	2.00	2.00	.00	.00	2.00	3.00	2.00	.00	2.00	.00
Range		3.00	3.00	3.00	3.00	3.00	3.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
Minimum		.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Maximum		3.00	3.00	3.00	3.00	3.00	3.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
Percentiles	25	1.00	1.00	1.00	1.00	1.00	1.00	2.00	.00	.00	2.00	2.00	2.00	.00	2.00	.00
	50	2.00	2.00	2.00	2.00	2.00	2.00	3.00	.00	.00	3.00	3.00	2.00	.00	3.00	2.00
	75	2.00	2.00	2.00	2.00	2.00	2.00	4.00	.75	2.00	4.00	3.00	3.00	2.00	3.00	3.00

		Intensity of Apd	Intensity of watery diarrhoea	Intensity of bloody mucoid diarrhoea	Intensity of tiredness	Intensity of nausea	Intensity of loss of appetite	Intensity of itching skin rash	Intensity of fever	Intensity of dizziness	Apd severity index based on frequency and intensity of symptom	Wdia severity index based on frequency and intensity of symptom	Bdia severity index based on frequency and intensity of symptom	Tired severity index based on frequency and intensity of symptom	Nos severity index based on frequency and intensity of symptom	App severity index based on frequency and intensity of symptom	Itch severity index based on frequency and intensity of symptom	Fever severity index based on frequency and intensity of symptom	Dizziness severity index based on frequency and intensity of symptom
N	Valid	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
	Missing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Median		2.50	.00	.00	3.00	2.00	2.00	.00	3.00	2.00	4.00	1.00	1.00	3.00	3.00	3.00	1.00	3.00	3.00
Mode		3.00	.00	.00	3.00	3.00	2.00	.00	3.00	.00	4.00	1.00	1.00	3.00	3.00	3.00	1.00	3.00	1.00
Range		4.00	3.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
Minimum		.00	.00	.00	.00	.00	.00	.00	.00	.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Maximum		4.00	3.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	5.00	4.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
Percentiles	25	2.00	.00	.00	2.00	1.00	1.00	.00	2.00	.00	3.00	1.00	1.00	3.00	2.25	2.25	1.00	3.00	1.00
	50	2.50	.00	.00	3.00	2.00	2.00	.00	3.00	2.00	4.00	1.00	1.00	3.00	3.00	3.00	1.00	3.00	3.00
	75	3.00	.75	1.00	3.00	3.00	3.00	2.00	3.00	3.00	4.00	1.75	2.00	4.00	4.00	4.00	3.00	4.00	4.00

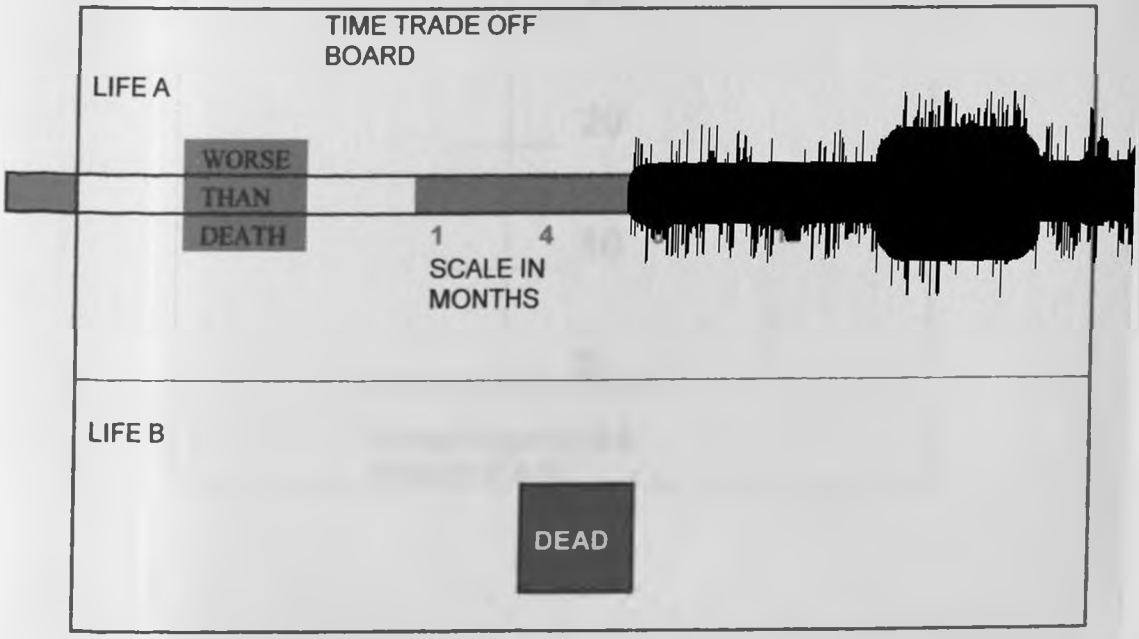
Key: Apd-abdominal pain and discomfort; Dia-diarrhoea; Wdia-watery diarrhoea; Bdia-bloody diarrhoea; Tired-tiredness; Nos-nausea; App-loss of appetite, Itch-itching skin rash

APPENDIX 7.3: PROPS: TTO BOARD, VAS FEELING THERMOMETER AND SG BOARD

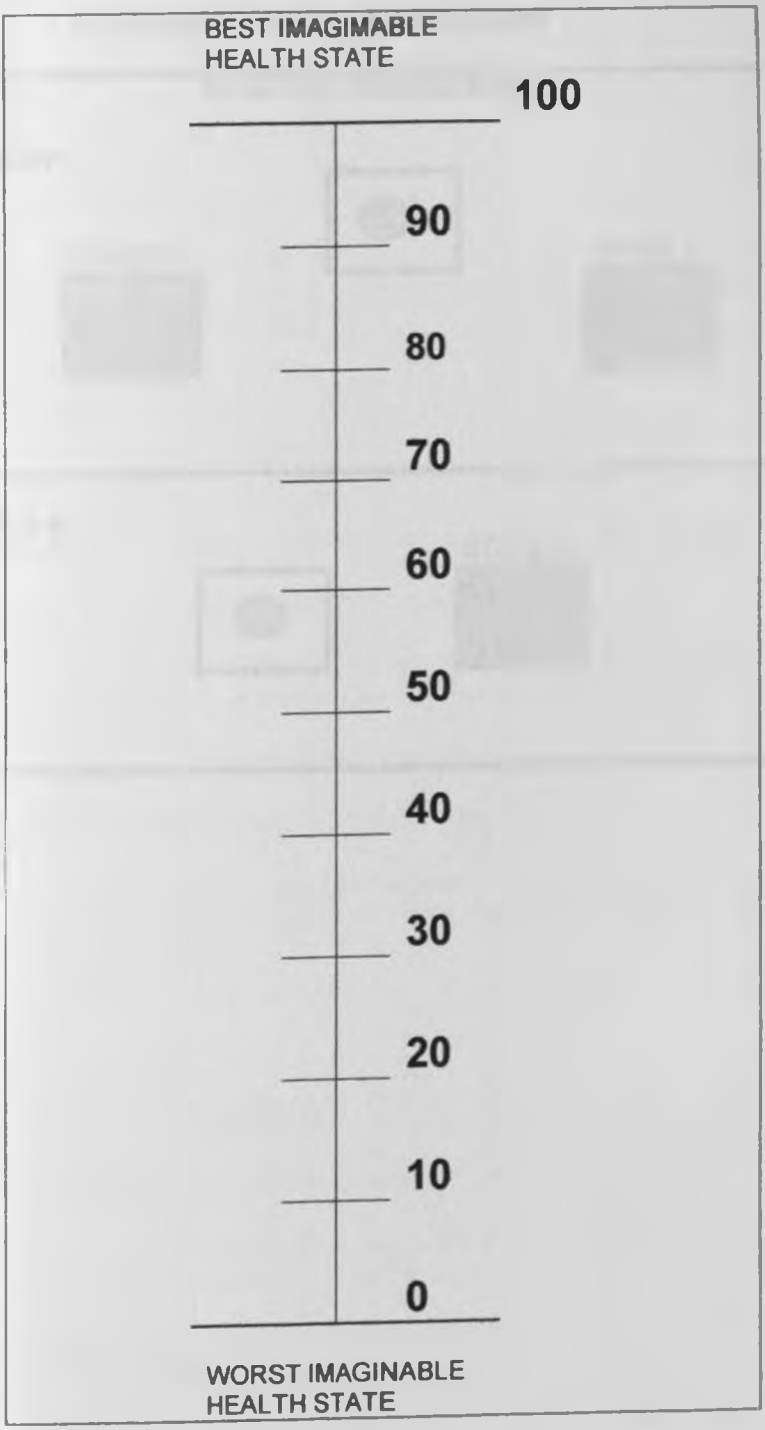
TIME TRADE OFF BOARD FOR STATES BETTER THAN DEATH



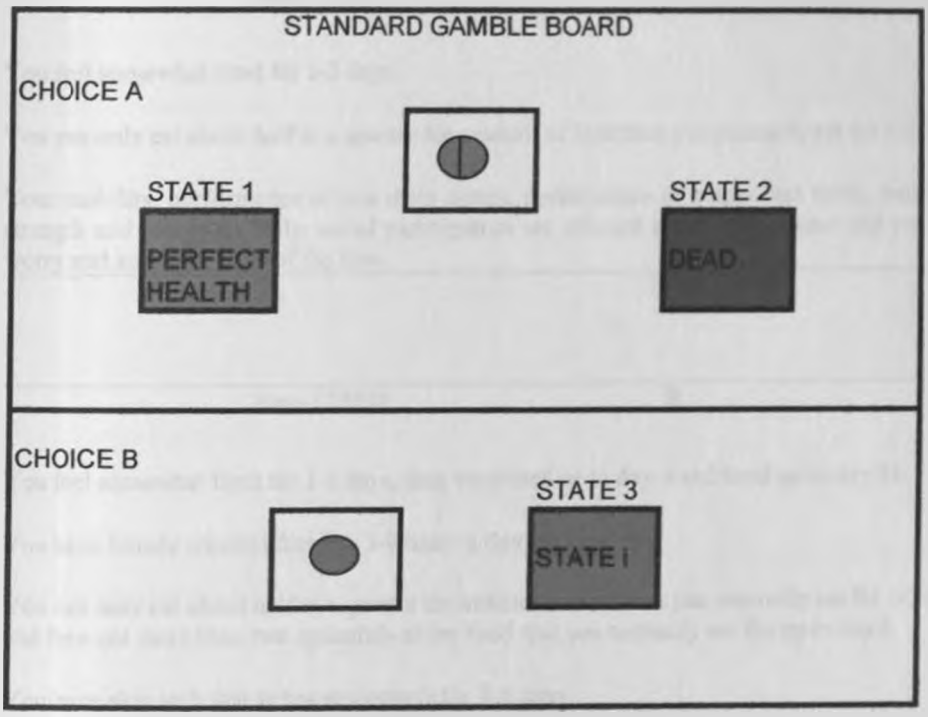
TIME TRADE OFF BOARD FOR STATES WORSE THAN DEATH



VISUAL ANALOG SCALE



STANDARD GAMBLE CHANCE BOARD



APPENDIX 7.4: SCHISTOSOMIASIS MANSONI DISEASE STATE SCENARIOS USED IN THE VALUATION STUDY

State 111111 A

- You feel somewhat tired for 1-3 days
- You can only eat about half to a quarter the amount of food that you normally eat for 1-3 days
- Your mobility, performance of you daily duties, performance of output and work, feeling of strength and energy in body, social participation are affected a little of the time and you feel worry and anxiety a little of the time.

State 122232 B

- You feel somewhat tired for 1-3 days, then very tired up to day 6 and tired up to day 14.
- You have bloody mucoid diarrhea 3-9 times a day for 1-6 days
- You can only eat about half to a quarter the amount of food that you normally eat for 1-3 days and then not more than two spoonfuls of the food that you normally eat for up to day 6.
- You have skin rash that itches moderately for 1-6 days
- Your mobility is affected is affected a little of the time
- Your performance of you daily duties, performance of output and work feeling of strength and energy in body and social participation are affected some of the time and you feel worry and anxiety some of the time.

State 213102 **C**

- You feel somewhat tired for 1-3 days and then very tired up to day 6.
- You can only eat half to a quarter of the food that you normally eat for 1-3 days and then no more than two spoonfuls the amount of food that you normally eat for up to day 6.
- Your mobility, performance of output and work are affected some of the time and you feel worry and anxiety some of the time.
- Your performance of you daily duties and feeling of strength and energy in body are affected a little of the time
- Your social participation is affected none of the time.

State 222222 **D**

- You feel very tired for 1-6 days
- You can only eat half to a quarter the amount of food that you normally eat for 1-3 days and no more than two spoonfuls of the food that you normally eat up to day 6
- Your mobility, performance of you daily duties, performance of output and work, feeling of strength and energy in body, social participation are affected some of the time and you feel worry and anxiety some of the time.

State 232222 **E**

- You feel very tired for up to 10 days
- You have watery diarrhea for less than a day
- You have bloody mucoid diarrhea 3-9 times a day for 1-3 days
- You can only eat half to a quarter of the food that you normally eat for 1-3 days and then no more than two spoonfuls of the food that you normally eat up to day 6
- You have skin rash that itches moderately for 1-6 days
- Your mobility, performance of you daily duties, performance of output and work, feeling of strength and energy in body, social participation are affected some of the time and you feel worry and anxiety some of the time.

State 333333 **F**

- You have watery diarrhea for less than a day
- You have bloody mucoid diarrhea 3-9 times a day for 1-3 days
- You feel extremely tired for 7-14 days
- You have skin rash that itches moderately for 1-6 days
- You can eat no more than two spoonfuls of the food that you normally eat for 1-6 days
- Your mobility, performance of you daily duties, performance of output and work, feeling of strength and energy in body, social participation are affected some of the time and you feel worry and anxiety some of the time.

Appendix 7.5: Aspects considered inappropriate in valuation instruments (n=16)

Attribute will:	VAS	TTO	SG
Annoy	• People may not take mentioning of death lightly as it may remind them of family members who have died (12.5%) • People thinking that they will be in states you ask them to value (12.5%) • If someone thinks you are asking about their physical and sexual maturity (6.25%)	• Mentioning death to those who fear death and the sickly. The old might think they have a short to live that's why you talk about death to them. Those suspicious of devil worshiping might think you will bring death to them. (25%) • Some may think you are teling them they will not live for long (12.5%)	• Misconception that the respondent would be in these states and therefore when asked to choose between a gamble (with immediate death) and certainty (bad state), it would be annoying because no one likes to die (25%).
Embarrass	• Some youth may mistake the local term for health (*ugima wa mwira*) to refer to sexual maturity (6.25%) • If one fails to understand the questions or cannot use numbers due to illiteracy (12.5%)		• Talking about symptoms like bloody diarrhoea by the young to the old. People don't like to talk about their illness or that in the family (18.75%)
Convey wrong meaning	• People thinking you are evaluating their own states (12.5%) • Exercise arousing suspicion in people about it's purpose (6.25%) • Mentioning death (6.25%)	• People misunderstanding the purpose of the exercise and wondering why you are talking about death (31.25%)	• Mentioning immediate death to an old or sick person may represent loss of hope in life and fasten death. If they have one symptom in the scenario, they may think they will get all the others. (18.75%). • Telling people they will be in perfect health is a lie (6.25%)
Cause problems		• Talk of death to the sickly as they might misunderstand and think they will die. (12.5%)	• If one chooses immediate death in a gamble and then they die by coincidence, the researcher might be accused if bewitching the person (6.25%)
Offend and upset		• The use of death as an endpoint might cause problems to some people (6.25%)	

Appendix 7.6: Worries, Feelings and thoughts provoked by use of techniques (n=16)

	VAS	TTO	SG
Things respondent worried about	• Thoughts about death (18.75%)	• Thinking about two lives and wondering how to avoid the problematic one (31.25%) • Asking about death (6.25%)	• "*Thinking that when I am gambling I can land in death, but one is also happy to know they can end up in perfect health*"..... "*Thinking about death and some very severe symptoms. If you imagine yourself in these you would feel worried.*" I thought those things might happen (18.75%).
Feelings thoughts or emotions provoked by use of instrument	• Thinking you are trying to value their health states. People might not like others to know where their states fall (18.75%), but I thought about where I would rank my health state. • It is unimaginable to think of death as 0 and perfect health as 100 (6.25%) • Thought about the future (6.25%).	• Mentioning death too often. It is not very good to think about death (12.5%) • Thinking of the good life in perfect health (6.25%)	• Thinking of the TTO and how I can avoid gambling (12.5%). • "*Death was an issue. I thought about the possibility of dying when the probability was high.*" The thought of dying and leaving so many things undone "*made me feel sad and worried.*" (18.75%)

APPENDIX 7.7: *PRE-TESTING INTERVIEW SCHEDULE*

HEALTH STATES VALUATION TECHNIQUES PRETEST QUESTIONNAIRE

[This questionnaire must be completed in pencil]

1	Questionnaire number		

		Day		Month		Year			
2	Date of interview.								

3. Name of interviewer_____

4. Name of respondent_____

5.

Location *[Circle one only]*	1. Tebere
	2. Mutithi
	3. Thiba
	4. Nyangiti
	5. Murinduko

6. Sub-location *[Circle one only]*

1	Kiarukungu	12	Wamumu
2	Mahigaini	13	Thiba
3	Gathigiriri	14	Mathangauta
4	Rukanga	15	Kirimara
5	Kinyaga	16	Ndomba
6	Mathigaini	17	Nyangati
7	Kombuini	18	Riagicheru
8	Kathiga	19	Mugabaciura
9	Kiandegwa	20	Miuu
10	Kabiriri	21	Kamunyange
11	Nguka		

VALUATION TECHNIQUES PRE-TEST

INTRODUCTION

I am *(say your name)* working with Mercy Mugo, a student researcher from the University of Nairobi. Last year between September and November, we learnt from some members of your community how bilharzia affects health and quality of life of people from your community. In the next few weeks, we want to visit some of the people we talked to in a follow up study to establish the value or worth of the states of health resulting from the interviews we did with them.

Before doing so, we would like to establish the relevance, appropriateness, meanings of words, thoughts and emotions that the questions we will ask provoke in people. This step is to help us design questions that people of Mwea understand using terms that they are familiar with and that have similar meanings, and are not offensive to anyone.

We are here today to ask you to help us to understand these terms and concepts. Can we go on? Thank you.

In the questions that we will ask members of the community, we will to use three methods to assign values to health states. For you to be able to answer the questions I have with me for you today, I will take you through a mock exercise with each of the methods to make sure that you understand the issues that I will be questioning you about. Note that in this exercise there are no right or wrong answers. **Your** views are most important to us. Feel free to say anything that comes to your mind when I ask the questions.

Since we are going to be talking and your views are very important to us, I would like to request you to allow me tape record our conversation, so that I can refer to it later on. Can I please use a tape recorder? Thank you. Now we shall start.

We shall start with a method called the VAS, then we do TTO and finally, SG. After each method I will ask you questions that will help us to decide upon the most appropriate method to use with people of Mwea.

Let me explain to you what we are going to do. First, we will use each of the methods, one at a time, to serve as an example. This is for you to understand how the method works and the questions asked in each method, as we will ask them to the members of the community. After the example, we will ask you questions about the method, in terms of the terms, ideas and concepts found in the method. We would urge you to think loudly and speak out your thoughts freely as we ask you these questions.

Can we start? Thank you.

VAS TECHNIQUE PRE-TEST

Explain the VAS task using the VAS script. Use the VAS example. Questions 7 and 8 to be answered by the interviewer.

7. Did the respondent understand the VAS task?

   1. Yes. 2. No.

8. How many times did you go through the VAS example? _____Times

Draw the respondent's attention to the following key words and phrases before asking the questions.

Key words and phrases

In the mock exercise that we have just completed, you have:

1. Ranked a health state
2. Expressed your preference for a heath states in terms of a number
3. Had to imagine yourself being in a health state scenario
4. Had to use the feeling thermometer to rank health states.

I would now like to ask you questions concerning these ideas or key words and phrases that we found in this method.

9. Tell me about some common instances and examples where people in this community may **rank** things.

10. Tell me about some common instances and examples where people in this community may **express their preference of something in terms of a number.**

11. Tell me about some common instances and examples of where you have **imagined** something or being in a situation.

12. Tell me the different ways that people in this community would talk about the term **health state.**

BEFORE ANSWERING QUESTIONS 13-18, I WANT YOU TO CONSIDER THE VAS TASK EXAMPLE THAT WE DID TOGETHER.

13. In your opinion, what aspects in these question might get someone annoyed and why?

14. In your opinion, what aspects in these question might get someone embarrassed and why?

15. Are there any parts of these questions that might convey the wrong meaning to someone and which ones? ***PROBE: What kinds of meanings are they are likely to convey?***

16. Are there any parts of these questions that might sound disrespectful to a person and which ones? ***PROBE: How would they be respectful?***

17. Are there any parts of these questions that might cause problems to anyone and which ones? ***PROBE: What kinds of problems?***

18. Think of all the things we did in that example and tell me the ones that might get someone upset or offended. PROBE: ***why would these things get someone upset or offended?***

I HAVE THESE TWO HEALH STATES WITH ME *(show heath state cards A and D).* I WILL READ OUT THEIR DESCRIPTIONS FOR YOU AND THEN ASK YOU TO ANSWER THE NEXT THREE QUUESTIONS. I WANT YOU TO IMAGINE YOURSELF BEING IN EACH OF THESE STATES OF HEALTH.*(Read out the description of each health state, letting the respondent know which is A and which is D. When you have read the two states ask the following questions)*

19. Which of these two health states would you say is better off and why?

20. What do you see as the differences between A and D?

21. Did you have any difficulties imagining and seeing yourself in these health states?
    1 yes ***(go to 22)*** 2. No. ***(go to 23)***

22. If yes, what was the nature of the difficulties?

23. Please tell me about any other difficulties you may have experienced in going through the VAS exercise

24. What is your opinion about being asked to place a numerical value on a health state?

25. When I asked you the VAS questions, did you think
    1. Very hard
    2. Hard
    3 A little
    4 very little?

26. Generally, did you find the VAS procedure,
    1. Very easy to understand
    2. Easy to understand
    3. Fairly easy to understand
    4. Difficult to understand
    5. Very difficult to understand?

27. Do you think it is bad to ask someone to imagine being dead or to think about death and why?

    _____

    _____

28. Do you think it is bad to ask someone to imagine being in a very sick state of health and why?

    _____

    _____

29. Do you think asking about death can cause one's death or death in this community and how?

    _____

    _____

30. Do you think asking about a sickly health state can cause sickness in this community and how?

    _____

    _____

31. Tell me about some of the things you worried about during the VAS exercise.

    _____

    _____

32. What was the strongest feeling, thought or emotion that you experienced as we went through the VAS exercise?

    _____

    _____

33. Can you think of all the different kinds of people who should not be asked these questions and why.

Kinds of people	Reasons they should not be asked

34. What would be the most polite way to ask these questions to these kinds of people?

Explain the TTO task using the TTO script. Use the TTO example. Questions 34 and 35 to be answered by the interviewer.

34. Did the respondent understand the TTO task?

    1. Yes. 2. No.

35. How many times did you go through the TTO example? _____Times

Draw the respondent's attention to the following key words and phrases before asking the questions.

Key words and phrases

In the mock exercise that we have just completed, you have:

1 given up time (months or years to remain healthy)
2 choose between health states (perfect health and less than perfect health)
3 imagined how being in a health state would be and imagined yourself being in these health states
4 had to consider future time in months or years (concept of the future and how far future is considered to be)
5 imagined your own death
6 imagined perfect health
7 traded off years in a poor state to attain perfect health
8 compared different lives
9 had to think about time
10 had to think of life in years

I would now like to ask you questions concerning these ideas or key words and phrases that we found in this method.

36. Tell me about some common instances and examples where people in this community make choices between different things and how they go about it.

_____

_____

37. Tell me about some common instances and examples where people in this community give up one thing for another and how they go about it.

38. Tell me about some common instances and examples about how people in this community talk about the future and what they consider future to be.

39. Tell me about some common instances and examples about how people in this community talk about perfect health and what they consider perfect health to be.

40. Tell me about how people in this community make comparisons between things. ***PROBE: Tell me the commonest forms of comparisons that people make.***

41. Tell me about how people in this community talk about time.

42. Tell me about how people in this community think and talk about life.

BEFORE ANSWERING QUESTIONS 43-48, I WANT YOU TO CONSIDER THE TTO EXAMPLE THAT WE DID TOGETHER.

43. In your opinion, what aspects in these question might get someone annoyed and why?

44. In your opinion, what aspects in these question might get someone embarrassed and why?

45. Are there any parts of these questions that might convey the wrong meaning to someone and which ones? ***PROBE: What kinds of meanings are they are likely to convey?***

46. Are there any parts of these questions that might sound disrespectful to a person and which ones? ***PROBE: How would they be respectful?***

47. Are there any parts of these questions that might cause problems to anyone and which ones? ***PROBE: What kinds of problems?***

48. Think of all the things we did in that example and tell me the ones that might get someone upset or offended. ***PROBE: why would these things get someone upset or offended?***

49. Please tell me about any difficulties you may have experienced in going through the TTO exercise

50. What is your opinion about being asked to give up some time in bad health so as to stay in good health?

51. When I asked you the TTO questions, did you think
    1. Very hard
    2. Hard
    3 A little
    4 very little?

52. Generally, did you find the TTO procedure

1 Very easy to understand
2 Easy to understand
3. Fairly easy to understand
4 Difficult to understand
5 Very difficult to understand ?

53. Tell me about some of the things you worried about during the TTO exercise.

54. What was the strongest feeling, thought or emotion that you experienced as we went through the TTO exercise?

55. Can you think of all the different kinds of people who should not be asked these questions and why.

Kinds of people	**Reasons they should not be asked**

56. What would be the most polite way to ask these questions to these kinds of people?

57. In using the TTO method, we touched on the subject of death. Please tell me your views about talking about death in this community?

PROMPT: ***what would be the best way to talk about death***

58. How do you feel about being asked to compare two lives and then choose one?

59. Normally, how far into the future do you consider? (***For example when making future plans)***

STANDARD GAMBLE PRE-TEST QUESTIONS

Explain the SG task using the SG script. Use the SG example. Questions 60 and 61 to be answered by the interviewer.

60. Did the respondent understand the SG task?

1. Yes. 2. No.

61. How many times did you go through the SG example? _____Times

Draw the respondent's attention to the following key words and phrases before asking the questions.

Key words and phrases

In the mock exercise that we have just completed, you:

1 Took a risk or chance in making a choice
2 Dealt with the concept of probability (and how it can be expressed)
3 Considered probability of immediate death
4 Gambled with states of health in making your choice
5 Considered and thought about death and immediate death
6 Dealt with changing chances or probability of being in states of health (how easy it is to conceptualize these)
7 dealt with the idea of chance board
8 uncertainty

I would now like to ask you questions concerning these ideas or key words and phrases that we found in this method.

62. Tell me about some common instances and examples where people in this community take risks and what they take into account in making such risks.

_____

_____

63. Tell me about some common instances and examples where people in this community would be involved in situations where they have to take a chance or probability of something.

_____

_____

64. Tell me about some common instances and examples where people in this community are involved in gambling.

65. Tell me about some common instances and examples where people in this community deal with uncertainty and how they express it.

BEFORE ANSWERING QUESTIONS 66-71, I WANT YOU TO CONSIDER THE SG EXAMPLE THAT WE DID TOGETHER.

66. In your opinion, what aspects in these question might get someone annoyed and why?

67. In your opinion, what aspects in these question might get someone embarrassed and why?

68. Are there any parts of these questions that might convey the wrong meaning to someone and which ones? ***PROBE: What kinds of meanings are they are likely to convey?***

69. Are there any parts of these questions that might sound disrespectful to a person and which ones? ***PROBE: How would they be respectful?***

70. Are there any parts of these questions that might cause problems to anyone and which ones? ***PROBE: What kinds of problems?***

71. Think of all the things we did in that example and tell me the ones that might get someone upset or offended. ***PROBE: why would these things get someone upset or offended?***

72. Please tell me about any difficulties you may have experienced in going through the SG exercise

73. What is your opinion about being asked to risk immediate death so as to stay in good health?

74. What is your opinion about being asked to gamble with being in perfect health or dying immediately and staying in a less tha perfect health state of life?

75. What do you feel when asked to take a chance of dying or staying in perfect health?

76. When I asked you the SG questions, did you think
    1. Very hard
    2. Hard
    3 A little
    4 very little?

77. Generally, did you find the SG procedure
    1 Very easy to understand
    2 Easy to understand
    3 Fairly easy to understand
    4 Difficult to understand
    5 Very difficult to understand ?

78. Tell me about some of the things you worried about during the SG exercise.

79. What was the strongest feeling, thought or emotion that you experienced as we went through the SG exercise?

80. Can you think of all the different kinds of people who should not be asked these questions and why.

Kinds of people	Reasons they should not be asked

81. What would be the most polite way to ask these questions to these kinds of people?

82. How do you feel about being asked to compare an uncertain life with a certain life and then choose one?

83. How comfortable are you dealing with the concept of probability?

84. Tell me about all the different ways of expressing chance, risk and probability in this community?

GENERAL QUESTIONS

(These questions should be asked to those respondents who complete two or three of the methods being pre-tested. Therefore the interviewer should rank either the two that apply or all the three)

85. Please rank the (two) three methods that we have considered in terms of which ones you would prefer to use in future if you were required to value health states. ***(Indicate preference where 1 is most preferred and (2) 3 least preferred)***

SG_____
TTO_____
VAS_____

86. What factors have you considered in your ranking above? ***(Can you please explain why you prefer 1 to 2 to 3).***

_____
_____
_____
_____

87. We would like to use the three methods with some members of this community to value several health states. If we take you to be a representative member of this community, which method would you recommend and why? ***(Where only two methods were pre-tested, ask the respondent which one they would recommend)***

_____
_____
_____
_____

To help in comparing responses from the various persons we have talked to I would like to ask you some general questions about yourself in this last section.

88 Name ***(optional)***_____

89. Sex: 1. Male 2. Female ***[circle one]***

90. What is your age? .....*(years).*

91. What is your current marital status? ***[circle the correct response]***

1. Single 2. Married 3. Co-habiting
4. Divorced / Separated 5. Widower/ Widow 6. Other (specify)

92. Do you have children? 1 Yes. 2. No.

93. If yes, how many child(ren)_____child(ren).

94. What is the highest level of education that you have reached? ***[Circle only one]***

	Educational level attained
1	None ***[Do not ask 135]***
2	Primary
3	Secondary ("O" level)
4	Secondary ("A" level)
5	Diploma
6	Graduate (degree)
7	Adult education
8	Other (specify)

95.

How many years did you spend in ***[Indicate number of years spent in school against each level up to the highest level reached as reported in 134 above]***	**Educational level attained**	**No. of years**
	Primary?	
	Secondary ("O" level)?	
	Secondary ("A" level)?	
	Diploma?	
	Graduate (degree)?	
	Adult education?	
	Other (specify)?	
Total years in school ***[Do not add during interview. Add up after]***		

96. What is your main profession?

1 Farmer
2 Teacher
3 Businessperson (specify type)
4 Casual laborer
5 Civil servant (specify)
6 Student
7 other (specify)

97. What do you do for a living?

1 Farmer
2 Teacher
3 Businessperson (Specify type)
4 Casual laborer
5 Civil servant (specify)
6 Student
7 Other (specify)

WE HAVE NOW COME TO THE END OF THIS INTERVIEW. I THANK YOU VERY MUCH FOR TAKING THE TIME TO HELP ME IN THIS EXERCISE.

Is there anything you would like to ask me before I leave? (Listen and answer honestly and to the best of your knowledge any questions the respondent may be having)

THANK YOU ONCE AGAIN.

SCHISTOSOMIASIS HEALTH STATES VALUATION SCRIPT

INTRODUCTION

I am (*say your name*) working with Mercy Mugo, a student researcher from the University of Nairobi. Last year between September and November, we learnt from you how bilharzia affects health and quality of life of people from your community. You accepted our request to visit you for a follow up study to establish the value or worth of the states of health that can result from having bilharzia.

We are here today to ask you to help us to establish the worth you attach to some of those health states. Can we go on? Thank you.

[If yes, present the information sheet and consent form and allow the respondent to read through. If unable to read, read out for her/him. Ask the respondent to sign the consent form, before continuing with the interview].

We are going to use two methods to give a value to each of the health states that I will present to you. Before we do each method, I will take you through an example to help you understand what each method requires and how it works. Note that in this exercise there are no right or wrong answers. It is **your** views that are most important to us.

Have the VAS and TTO props ready. Keep health state cards PH, A, C, E, F and ZZ on standby. Remove the VAS prop, the pointer and the mock health state card MM.

WE ARE NOW GOING TO START WITH THE FIRST METHOD CALLED THE VISUAL ANALOG SCALE.

A: VISUAL ANALOG SCALE HEALTH STATES RANKING

(*Show respondent the visual analog scale*). **THIS IS A SCALE. IT IS KNOWN AS A FEELING THERMOMETER SCALE. WE CALL IT A FEELING THEREMOMETER BECAUSE IT IS USED TO MEASURE PEOPLE'S FEELINGS ABOUT DIFFERENT HEALTH STATES. WE ARE GOING TO USE IT TO MEASURE HOW GOOD OR BAD YOU FEEL A HEALTH STATE IS BY GIVING IT A NUMBER.**

(*Put the VAS prop on a stable place*). **THIS IS HOW IT WORKS. AS YOU CAN SEE, THE BOTTOM OF THE SCALE HAS 0 AND THE TOP HAS 100. THE MORE PREFFERABLE OR DESIREABLE YOU FEEL A HEALTH STATE IS, THE CLOSER IT SHOULD BE TO THE TOP AND THE LESS PREFFERABLE OR DESIRABLE YOU FEEL A HEALTH STATE IS THE CLOSER IT SHOULD BE TO THE BOTTOM.**

EXAMPLE

I WILL TAKE YOU THROUGH AN EXAMPLE, AND THEN I WILL ASK YOU TO VALUE 6 STATES OF HEALTH THAT I WILL PRESENT TO YOU. (*Remove the mock health state card M, and say to the respondent).* ON THIS CARD IS A DESCRIPTION OF A HEALTH STATE WHICH I AM NOW GOING TO READ OUT FOR YOU. IT SAYS...*(Read out the health state description on the card).* NOW I WANT YOU TO IMAGINE THIS PERSON WHO IS JUST LIKE YOU BEING IN THE STATE OF HEALTH DESCRIBED IN THIS CARD. CONSIDER THAT THE TOP OF THIS SCALE *(Point to the top of VAS)* INDICATES 100 AS THE MOST DESIRABLE HEALTH STATE, AND THE BOTTOM *(Point to the bottom of VAS)* INDICATES 0 AS THE LEAST DESIRABLE HEALTH STATE. USING THIS POINTER *(Give one pointer to the respondent)* SHOW ME WHERE YOU WOULD PLACE THIS HEALTH STATE BETWEEN THESE TWO ENDS.

(PROMPT: ***Do you understand how this method works? Can you tell me how you have understood this method works? If they do not understand, repeat this example until they understand. Note the number of times you repeat and record in the record form question 10, the number of times you explained the procedure. If yes say ….).***

WELL DONE. NOW THAT YOU UNDERSTAND THE VAS PROCEDURE, LET US GO AHEAD AND FIND OUT WHAT VALUES YOU FEEL THE FOLLOWING 6 HEALTH STATE SCENARIOS SHOULD BE ASSIGNED.

I AM NOW GOING SHOW YOU THE DESCRIPTIONS OF 6 HEALTH STATE SCENARIOS. ***(Remove health state scenario cards PH, A, C, E, F, AND ZZ. Show the 6 health states cards).*** **EACH OF THESE CARDS CONTAINS A DESCRIPTION OF A HEALTH STATE SCENARIO. EACH HEALTH STATE SCENARIO IS DESCRIBED IN TERMS OF DIFFERENT SYMPTOMS OF BILHARZIA AND HOW OFTEN SOME 3 HEALTH STATE DOMAINS ARE AFFECTED. THE THREE HEALTH STATE DOMAINS ARE: PERFORMANCE OF WORK AND DAILY DUTIES (i.e. playing, farming, household chores, usual activity done for livelihood, school work etc.); ABILITY TO BE IN SOCIAL FUNCTIONS (i.e. group meetings, religious meetings, weddings, visiting friends and family and socializing with friends and colleagues); and FEELING WORRIED AND ANXIOUS.**

EACH CARD IS INDENTIFIED BY A NAME OF A PERSON AND A LETTER (*show names and letters PH, A, C, E, F, AND ZZ)* TO SHOW THAT THE HEALTH STATE SCENARIOS ARE DIFFERENT. I WILL CALL EACH CARD BY THE NAME OF THE IMAGINARY PERSON. ***(Call a male or female name depending on the gender of the respondent)*** **I WANT YOU TO THINK OF THE HEALTH STATE THIS PERSON IS IN AND THEN SHOW ME HOW DESIRABLE IT WOULD BE TO YOU BY PLACING IT ON THIS SCALE. REMEMBER, IT IS NOT YOUR HEALTH STATE WE ARE VALUING BUT THE HEALTH STATES OF THESE IMAGINERY PERSONS DESCRIBED BY EACH CARD.**

NOW, I WILL READ OUT FOR YOU EACH OF THE HEALTH STATE SCENARIO ONE AT A TIME. USING THIS POINTER (*Show the pointer*) I WILL THEN ASK YOU TO POINT WHERE ON THIS FEELING THERMOMETER, BETWEEN 0 AND 100 (*Point to 0 and 100 on the VAS scale*) YOU FEEL THE SCENARIO SHOULD BE PLACED TO SHOW HOW GOOD OR BAD YOU FEEL A HEALTH STATE IS. FEEL FREE TO CHANGE THE ORDER UNTIL YOU ARE SATISFIED WITH YOUR RANKING OF THE HEALTH STATE SCENARIOS FROM LEAST DESIRABLE TO MOST DESIRABLE.

(Have the scoring sheet for VAS ready.) **LET US NOW BEGIN. (*Record the start time in the record form question 11*). I WILL START WITH THIS HEALTH STATE SCENARIO LABELLED PH.** ***(Read out each health state scenario at a time, slowly starting with PH. Allow the respondent time to visualize and understand the contents of the health state scenario. Ask them to imagine the person named on the card being in this health state.)*** **I WANT YOU TO THINK OF THIS PERSON (*Mention the name on the card, either male or female depending on the gender of the respondent)* TO BE JUST LIKE YOU.** ***Then ask him or her to place it on the feeling thermometer between 0 and 100. Record the value on the score sheet. Repeat this procedure for state ZZ. Finally, repeat the procedure for the next 4 states in this order; A, C, E and F.***

Should the respondent wish to change his or her previous ranking, allow this and record the new value accordingly. In cases where the respondent assigns the state being valued a value already assigned to a previously valued state, read out the description of the previous state and ask the respondent, **"DO YOU STILL WISH TO GIVE THESE TWO STATES THE SAME VALUE?"** *If the response says yes, record the value as it is. If the answer is no, adjust the values of the relevant states as per the respondents valuations.*

Then record the final values on the score sheet provided on the record form.

When the value of each of the states has been established and the respondent is satisfied with his or her ranking, complete the visual analog scale ranking scoring sheet on the record form, by transferring the health states as the respondent arranged them on the large portable VAS FEELING THERMOMETER by drawing a line form the state to the VAS scale. Make sure the arrows from the health states to feeling thermometer clearly point the respondent's value when transferred to the VAS on the record form.

Record the time this exercise ends on question 12 in the record form.

THANK YOU. WE HAVE FINISHED THAT PART OF THE EXERCISE. LET US NOW GO TO THE NEXT METHOD

(Return the VAS prop and pointers to the carry bag. Arrange the health states cards in order; PH, A, F, C, E, ZZ. Move on to TTO exercise.

B: TIME TRADE OFF HEALTH STATES VALUATION

B1: TTO PROCESS

THIS METHOD IS DIFFERENT FROM THE ONE WE HAVE JUST CONCLUDED. IT IS ABOUT HOW MUCH TIME YOU WOULD BE WILLING TO GIVE UP TO REMAIN IN A PERFECTLY HEALTHY STATE.

THIS IS HOW IT WORKS. WE WILL USE THE SAME HEALTH STATES THAT YOU HAVE JUST RANKED. I WILL PRESENT YOU WITH TWO HEALTH STATES. ONE OF THEM IS PERFECT HEALTH. THE OTHER IS THE STATE WE WANT TO GIVE A VALUE. I WILL PRESENT YOU THE TWO STATES AND ASK YOU TO TELL ME THE ONE YOU PREFER. IF YOU THINK THE TWO CHOICES ARE THE SAME, TELL ME.

TO MAKE THE EXERCISE EASIER TO UNDERSTAND, WE WILL USE AN AID. THIS AID IS CALLED THE TIME TRADE OFF BOARD. (*Show the time trade off board*) AS YOU CAN SEE, THE TOP PART OF THE BOARD IS LABBELED LIFE A AND THE LOWER PART IS LABELLED LIFE B. THESE ARE THE TWO CHOICES, BETWEEN LIFE A AND LIFE B.

EACH LIFE IS DESCRIBED IN TERMS OF DIFFERENT SYMPTOMS AND HOW OFTEN SOME 3 HEALTH STATE DOMAINS ARE AFFECTED. THE THREE HEALTH STATE DOMAINS ARE: PERFORMANCE OF WORK AND DAILY DUTIES (i.e. playing, farming, household chores, usual activity done for livelihood, school work etc.); ABILITY TO BE IN SOCIAL FUNCTIONS (i.e. group meetings, religious meetings, weddings, visiting friends and family and socializing with friends and colleagues); and FEELING WORRIED AND ANXIOUS.

THE DESCRIPTIONS OF EACH LIFE SCENARIO DIFFER AS SHOWN ON THESE CARDS (*Show health state scenario cards to the respondent again and remind him they are the same you used earlier*). WHILE MAKING THE CHOICE, I WILL PUT A CARD ON EACH OF THESE POCKETS (Point to the pocket for life A and life B). I WILL THEN CHANGE THE TIME IN LIFE A WHILE LIFE B REMAINS THE SAME AND ASK YOU TO TELL ME WHICH YOU PREFER, A OR B, OR WHETHER YOU THINK THEY ARE THE SAME. *(Give the example of a weighing scale to remind the respondent that you are trying to establish the value of state in life B, as follows…).*

THINK OF THIS TTO BOARD AS A WEIGHING MACHINE. ON A WEIGHING MACHINE WE PLACE, SAY, THE AMOUNT OF RICE WHOSE WEIGHT WE WANT TO KNOW ON ONE SIDE. THEN ON THE OTHER WE KEEP CHANGING THE WEIGHTS (STONES) UNTIL THE TWO SIDES ARE BALANCING, THEN WE KNOW THE WEIGHT OF THE RICE. IN THE SAME WAY, WE WANT TO KNOW THE WORTH YOU GIVE THE HEALTH STATE IN LIFE B. SO WE WILL KEEP CHANGING THE TIME IN LIFE A UPTO THE POINT WHERE YOU FEEL THAT THEY ARE BALANCED. SOMETIMES LIFE A WILL BE BETTER AND AT OTHER TIMES LIFE B WILL BE BETTER, SO WE WILL KEEP CHANGING UNTIL YOU FEEL THEY ARE THE SAME.

REMEMBER THAT IT IS NOT YOUR HEALTH STATE THAT WE ARE VALUING. IT IS YOU WHO IS GOING TO TELL US THE VALUE YOU CAN GIVE TO THESE HEALTH STATES THAT WE HAVE DESCRIBED AND INDENTIFIED BY IMAGINARY NAMES AND LETTERS. THINK OF THE PEOPLE IN THESE STATES AS BEING JUST LIKE YOU.

EXAMPLE

LET US GO THROUGH AN EXAMPLE TOGETHER. ***(Set the TTO board on a stable surface. Remove the mock health state card (card MM) and the perfect health state card (Card PH)). Read out the contents of each card and make sure the respondent understands what health state scenario is contained in the card. Place card PH in pocket named life A and place M in pocket named life B. Set life A and life B at 12 months).***

AS YOU CAN SEE (*Point to both scales*) EACH SCALE HAS 12 MONTHS. THE PINK COLOUR ON THE TOP SCALE INDICATES HEALTHY LIFE. THE GREEN COLOUR ON THE BOTTOM SCALE INDICATES A LIFE WITH SOME HEALTH PROBLEMS. I WANT YOU TO IMAGINE *(mention name on the card in life A)* AND (*mention name on the card in life B*) IN THESE STATES. ***(Set both life A scale and life B scale to 12 months and pointing to the scales say).***

B1a. Each scale says 12 months. This means that you would either live in life A for 12 months or live in life B for 12 months. What happens after the 12 months is not known. Would you prefer A or B or are they equal?

(PROMPT: ***Do you understand how this method works? Can you tell me how you have understood this method works? If they do not understand, repeat this example until they understand. Note the number of times you explain and answer questions 19 and 21 in the TTO record form.)***

WELL DONE. NOW LET US GO ON AND VALUE THE FOLLOWING FOUR HEALTH STATE SCENARIOS.

B: TIME TRADE OFF HEALTH STATES VALUATION

B1: TTO PROCESS

THIS METHOD IS DIFFERENT FROM THE ONE WE HAVE JUST CONCLUDED. IT IS ABOUT HOW MUCH TIME YOU WOULD BE WILLING TO GIVE UP TO REMAIN IN A PERFECTLY HEALTHY STATE.

THIS IS HOW IT WORKS. WE WILL USE THE SAME HEALTH STATES THAT YOU HAVE JUST RANKED. I WILL PRESENT YOU WITH TWO HEALTH STATES. ONE OF THEM IS PERFECT HEALTH. THE OTHER IS THE STATE WE WANT TO GIVE A VALUE. I WILL PRESENT YOU THE TWO STATES AND ASK YOU TO TELL ME THE ONE YOU PREFER. IF YOU THINK THE TWO CHOICES ARE THE SAME, TELL ME.

TO MAKE THE EXERCISE EASIER TO UNDERSTAND, WE WILL USE AN AID. THIS AID IS CALLED THE TIME TRADE OFF BOARD. (*Show the time trade off board*) AS YOU CAN SEE, THE TOP PART OF THE BOARD IS LABBELED LIFE A AND THE LOWER PART IS LABELLED LIFE B. THESE ARE THE TWO CHOICES, BETWEEN LIFE A AND LIFE B.

EACH LIFE IS DESCRIBED IN TERMS OF DIFFERENT SYMPTOMS AND HOW OFTEN SOME 3 HEALTH STATE DOMAINS ARE AFFECTED. THE THREE HEALTH STATE DOMAINS ARE: PERFORMANCE OF WORK AND DAILY DUTIES (i.e. playing, farming, household chores, usual activity done for livelihood, school work etc.); ABILITY TO BE IN SOCIAL FUNCTIONS (i.e. group meetings, religious meetings, weddings, visiting friends and family and socializing with friends and colleagues); and FEELING WORRIED AND ANXIOUS.

THE DESCRIPTIONS OF EACH LIFE SCENARIO DIFFER AS SHOWN ON THESE CARDS (*Show health state scenario cards to the respondent again and remind him they are the same you used earlier*). WHILE MAKING THE CHOICE, I WILL PUT A CARD ON EACH OF THESE POCKETS (Point to the pocket for life A and life B). I WILL THEN CHANGE THE TIME IN LIFE A WHILE LIFE B REMAINS THE SAME AND ASK YOU TO TELL ME WHICH YOU PREFER, A OR B, OR WHETHER YOU THINK THEY ARE THE SAME. *(Give the example of a weighing scale to remind the respondent that you are trying to establish the value of state in life B, as follows…).*

THINK OF THIS TTO BOARD AS A WEIGHING MACHINE. ON A WEIGHING MACHINE WE PLACE, SAY, THE AMOUNT OF RICE WHOSE WEIGHT WE WANT TO KNOW ON ONE SIDE. THEN ON THE OTHER WE KEEP CHANGING THE WEIGHTS (STONES) UNTIL THE TWO SIDES ARE BALANCING, THEN WE KNOW THE WEIGHT OF THE RICE. IN THE SAME WAY, WE WANT TO KNOW THE WORTH YOU GIVE THE HEALTH STATE IN LIFE B. SO WE WILL KEEP CHANGING THE TIME IN LIFE A UPTO THE POINT WHERE YOU FEEL THAT THEY ARE BALANCED. SOMETIMES LIFE A WILL BE BETTER AND AT OTHER TIMES LIFE B WILL BE BETTER, SO WE WILL KEEP CHANGING UNTIL YOU FEEL THEY ARE THE SAME.

REMEMBER THAT IT IS NOT YOUR HEALTH STATE THAT WE ARE VALUING. IT IS YOU WHO IS GOING TO TELL US THE VALUE YOU CAN GIVE TO THESE HEALTH STATES THAT WE HAVE DESCRIBED AND INDENTIFIED BY IMAGINARY NAMES AND LETTERS. THINK OF THE PEOPLE IN THESE STATES AS BEING JUST LIKE YOU.

EXAMPLE

LET US GO THROUGH AN EXAMPLE TOGETHER. ***(Set the TTO board on a stable surface. Remove the mock health state card (card MM) and the perfect health state card (Card PH)). Read out the contents of each card and make sure the respondent understands what health state scenario is contained in the card. Place card PH in pocket named life A and place M in pocket named life B. Set life A and life B at 12 months).***

AS YOU CAN SEE (*Point to both scales*) EACH SCALE HAS 12 MONTHS. THE PINK COLOUR ON THE TOP SCALE INDICATES HEALTHY LIFE. THE GREEN COLOUR ON THE BOTTOM SCALE INDICATES A LIFE WITH SOME HEALTH PROBLEMS. I WANT YOU TO IMAGINE *(mention name on the card in life A)* AND (*mention name on the card in life B*) IN THESE STATES. ***(Set both life A scale and life B scale to 12 months and pointing to the scales say).***

B1a. Each scale says 12 months. This means that you would either live in life A for 12 months or live in life B for 12 months. What happens after the 12 months is not known. Would you prefer A or B or are they equal?

(PROMPT: ***Do you understand how this method works? Can you tell me how you have understood this method works? If they do not understand, repeat this example until they understand. Note the number of times you explain and answer questions 19 and 21 in the TTO record form.)***

WELL DONE. NOW LET US GO ON AND VALUE THE FOLLOWING FOUR HEALTH STATE SCENARIOS.

B2: TIME TRADE OFF INTERVIEW

(Begin the TTO interview by using the script for states better than death. On the record scripts provided, record the respondent's choice by indicating with a tick either choice A or choice B, or whether they are equal in the appropriate columns. If you get two consecutive ticks in the first two rows of column A, switch to script for states worse than death. Otherwise, continue the interview up to the point where you get two consecutive ticks in either column A or B or where the respondent indicates the two choices are equal. Then remove the next health state scenario to be valued and repeat the procedure. This later condition applies to valuations for states considered worse than death.)

(THIS SCRIPT SHOULD BE USED FOR STATES BETTER THAN DEATH)

Have the 6 health states cards ready arranged in this order, PH, A, F, C, E, ZZ.

Set the TTO board on a stable surface

Set life A at 12 months and life B at 12 months

Before inserting any health state card into the pockets, read out it's contents to the respondent carefully and slowly, making sure that they understand the health state scenario contained in the card. The health states should be valued in the following order. A, F, C, E.

[INFORMATION TO INTERVIEWERS: A separate sheet is provided for each health state scenario in the record form. Make sure that you have the appropriate sheet for the scenario in question ready before you begin valuing. Record the respondent's choice by indicating with a tick either choice A or choice B, or whether they are equal in the appropriate columns. If you get two consecutive ticks in the first two rows of column A, switch to script for states worse than death. Otherwise, continue the interview up to the point where you get two consecutive ticks in either column A or B or where the respondent indicates the two choices are equal. Then remove the next health state scenario to be valued and repeat the procedure.]

Record time TTO exercise starts in the TTO record form question 23. **WE ARE NOW GOING TO START OUR VALUATION EXERCISE BY VALUING HEALTH STATE SCENARIO A. SINCE I AM GOING TO BE ASKING YOU TO MAKE A CHOICE BETWEEN LIFE A AND LIFE B, LET ME START BY READING FOR YOU THE DESCRIPTION OF LIFE A** ***(Read out the description contained in card PH).*** **I HAVE NOW PLACED THIS CARD IN LIFE A POCKET. LIFE B IS DESCRIBED AS FOLLOWS,** ***(Read out the description contained in card A).*** **I HAVE NOW PLACED THIS CARD IN LIFE B POCKET.** ***(Start from question 1 and proceed as instructed in the paragraph under B2. Repeat the procedure with the next 3 health state scenarios to be valued, in the prescribed order. When all the 4 health state scenarios have been valued, record the time the interview ends in the TTO record form question 30.)*** **AS YOU CAN SEE...** ***(Go to 1).***

1. Each scale says 12 months. This means that you would either live in life A like (mention name of person on pink card) for 12 months (***point to the pink card***) or live in life B like (***mention name of person on the appropriate colored card***) for 12 months (point to the appropriate colored card. What happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you tick B or =, ask "why do you chose a worse off health state over perfect health?" Answer question 6 under each health state scenario. Remove the next state to be valued and start from 1. Otherwise ask 2.***

Move life A slider to 0, then ask 2.

2. Now I have changed life A scale to 0 month (*remove the pink card from pocket*). Life B scale remains unchanged at 12 months (***point to the appropriate colored card***). This means that you would either die immediately or live in life B like (***mention name of person on the appropriate colored card***) for 12 months and what happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you get two consecutive ticks under A, switch to script for states considered worse than death, otherwise proceed to the next question until you get two consecutive ticks in the same column or = in the middle column. Stop and remove the next scenario to be valued and start from 1. Otherwise ask 3.***

Move life A slider to 11 months, then ask 3

3. Now I have changed life A scale to 11 months (***point to the pink card***). Life B scale remains unchanged at 12 months (***point to the appropriate colored card)***. This means that you would either live in life A like (***mention name of person on pink card***) for 11 months or live in life B like (***mention name of person on the appropriate colored card***) for 12 months. What happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in the same column or = in the middle column, stop. Remove the next scenario to be valued and start from 1. Otherwise ask 4)***

Move life a slider to 1 month and ask 4.

4. Now I have changed life A scale to 1month (***point to the pink card***). Life B scale remains unchanged at 12 months (***point to the appropriate colored card***). This means that you would either live in life A like (***mention name of person on pink card***) for 1 month or live in life B like (***mention name of person on the appropriate colored card***) for 12 months. What happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in the same column or = in the middle column, stop. Remove the next scenario to be valued and start from 1. Otherwise ask 5)***

Move life a slider to 10 month and ask 5

5. Now I have changed life A scale to 10 months (***point to the pink card***). Life B scale remains unchanged at 12 months (***point to the appropriate colored card***). This means that you would either live in life A like ***(mention name of person on pink card***) for 10 months or live in life B like (***mention name of person on the appropriate colored card***) for 12 months. What happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in the same column or = in the middle column, stop. Remove the next scenario to be valued and start from 1. Otherwise ask 6)***

Move life a slider to 2 month and ask 6

6. Now I have changed life A scale to 2 months (***point to the pink card***). Life B scale remains unchanged at 12 months (***point to the appropriate colored card***). This means that you would either live in life A like (***mention name of person on pink card***) for 2 months or live in life B like (***mention name of person on the appropriate colored card***) for 12 months. What happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in the same column or = in the middle column, stop. Remove the next scenario to be valued and start from 1. Otherwise ask 7)***

Move life a slider to 9 month and ask 7.

7. Now I have changed life A scale to 9 months (***point to the pink card***). Life B scale remains unchanged at 12 months ***(point to the appropriate colored card***). This means that you would either live in life A like (***mention name of person on pink card***) for 9 months or live in life B like (***mention name of person on the appropriate colored card***) for 12 months. What happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in the same column or = in the middle column, stop. Remove the next scenario to be valued and start from 1. Otherwise ask 8)***

Move life a slider to 3month and ask 8.

8. Now I have changed life A scale to 3 months (***point to the pink card***). Life B scale remains unchanged at 12 months (***point to the appropriate colored card***). This means that you would either live in life A like (***mention name of person on pink card***) for 3 months or live in life B like (***mention name of person on the appropriate colored card***) for 12 months. What happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in the same column or = in the middle column, stop. Remove the next scenario to be valued and start from 1. Otherwise ask 9)***

Move life a slider to 8 month and ask 9

9. Now I have changed life A scale to 8 months (***point to the pink card***). Life B scale remains unchanged at 12 months (***point to the appropriate colored card***). This means that you would either live in life A like (***mention name of person on pink card***) for 8 months or live in life B like (***mention name of person on the appropriate colored card***) for 12 months. What happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in the same column or = in the middle column, stop. Remove the next scenario to be valued and start from 1. Otherwise ask 10)***

Move life a slider to 4 months and ask 10

10. Now I have changed life A scale to 4 months (***point to the pink card***). Life B scale remains unchanged at 12 months (***point to the appropriate colored card***). This means that you would either live in life A like ***(mention name of person on pink card***) for 4 months or live in life B like (***mention name of person on the appropriate colored card***) for 12 months. What happens after is not known. Would you prefer A or B or are they equal? ***(Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in the same column or = in the middle column, stop. Remove the next scenario to be valued and start from 1. Otherwise ask 11)***

Move life a slider to 7 months and ask 11.

11. Now I have changed life A scale to 7 months ***(point to the pink card)***. Life B scale remains unchanged at 12 months (***point to the appropriate colored card***). This means that you would either live in life A like ***(mention name of person on pink card)*** for 7 months or live in life B like (***mention name of person on the appropriate colored card***) for 12 months. What happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in the same column or = in the middle column, stop. Remove the next scenario to be valued and start from 1. Otherwise ask 12)***

Move life a slider to 5 months and ask 12.

12. Now I have changed life A scale to 5 months (***point to the pink card***). Life B scale remains unchanged at 12 months (***point to the appropriate colored card***). This means that you would either live in life A like (***mention name of person on pink card***) for 5 months or live in life B like (***mention name of person on the appropriate colored card***) for 12 months. What happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in the same column or = in the middle column, stop. Remove the next scenario to be valued and start from 1. Otherwise ask 13)***

Move life a slider to 6 month and ask 13.

13. Now I have changed life A scale to 6 months (***point to the pink card***). Life B scale remains unchanged at 12 months (***point to the appropriate colored card***). This means that you would either live in life A like (***mention name of person on pink card***) for 6 months or live in life B like (***mention name of person on the appropriate colored card***) for 12 months. What happens after is not known. Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in the same column or = in the middle column, stop. Remove the next scenario to be valued and start from 1. Otherwise ask 14)***

Ask 14.

14. Tell me the amount of time in weeks you are willing to give up in perfect health to avoid living in life B? _____Weeks.

Record time TTO exercise ends in question 30 in the record form

When you have finished valuing the six health states, say .....

THANK YOU. WE HAVE FINISHED WITH THAT METHOD.

B3: TIME TRADE OFF INTERVIEW

(*THIS SCRIPT SHOULD BE USED FOR STATES WORSE THAN DEATH*)

PROCEDURE FOR PRESENTING TTO CHOICES FOR STATES WORSE THAN DEATH

YOUR CHOICE HERE (*pointing to the TTO board*) INDICATES THAT YOU WOULD RATHER DIE IMMEDIATELY THAN LIVE IN THIS HEALTH STATE. TO VALUE THIS STATE WE ARE NOW GOING TO USE ANOTHER BOARD FOR PRESENTING CHOICES FOR STATES WORSE THAN DEATH. THIS IS HOW IT WORKS.

I HAVE NOW CHANGED THE CHOICES. AS BEFORE YOU WILL CHOSE BETWEEN A AND B. HOWEVER, B IS NOW IMMEDIATE DEATH (*put immediate death card in life B pocket*) WHILE IN A YOU WILL LIVE IN TWO STATES. THE FIRST STATE IS THE ONE YOU CONSIDERED WORSE THAN DEATH, FOLLOWED BY A SECOND STATE WHICH IS PERFECT HEALTH AND THEN WHAT HAPPENS AFTER IS NOT KNOWN. I WILL KEEP CHANGING THE TIME IN THE TWO STATES IN LIFE A, AND THEN ASK YOU TO CHOSE EITHER A OR B, OR TELL ME IF THEY ARE THE SAME. NOTE THAT TIME IN PERFECT HEALTH (PINK) AND TIME IN STATE BEING VALUED (GREEN), WHICH CONSTITUTE LIFE A, ADD UP TO 12 MONTHS AND WHAT HAPPENS AFTRE THE 2 MONTHS IS NOT KNOWN.

LET ME TAKE YOU THROUGH AN EXAMPLE

EXAMPLE

(Turn over the board to the TTO board for states worse than death. Place the perfect health card on the left and the mock health state card M1 on the right. Indicate that the pink color shows time in perfect health and the green color time in the health state being valued, that is considered worse than death. Point out to the respondent that this constitutes life A choice. Place immediate death card on life B pocket. Point out to the respondent that the blue color indicates immediate death. Set life A slider at 6 months so that pink and green occupy 6 months each.

AS YOU CAN SEE *(point to the scale on life A)* THIS SCALE SHOWS THAT (*mention name of he person on the card*) WILL LIVE FOR 6 MONTHS IN THIS STATE *(state being valued)* AND THEN RETURN TO PERFECT HEALTH

FOR 6 MONTHS. WHAT HAPPENS AFTER IS NOT KNOWN. CHOICE B SHOWS IMMEDIATE DEATH. WOULD YOU PREFER LIFE A OR B, OR WOULD YOU SAY THEY ARE THE SAME?

(PROMPT: ***Do you understand how this method works? Can you tell me how you have understood this method works? If they do not understand, repeat this example until they understand. Note the number of times you explain and answer questions 20 and 22 in the TTO record form.*****)**

WELL DONE. NOW LET US GO ON AND VALUE THIS HEALTH STATE SCENARIO.

Set the slide for life A at 0 month in the state being valued and 12 months in perfect health. Showing the respondent say:

1. Life A now is 0 months in this state (***point to the appropriate colored card and say like… mention name of person on card***) followed by 12 months in perfect health (***point pink card and say like ….mention name of person on card***). What happens after is not known. Life B is immediate death (***like….mention name of person on card***). Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you tick under B or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1 in the script for states better than death. If you tick under A, ask 2)***

   Move slider to 11 months in state 1 and 1 months in state 2. Ask 2

2. Life A is now 11 months in this state (***point to the appropriate colored card and say like… mention name of person on card***) followed by 1 months in perfect health (***point pink card and say like ….mention name of person on card***). What happens after is not known. Life B is immediate death (***like….mention name of person on card***). Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in one column or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1. Otherwise, ask 3***

   Move slider to 1 months in state 1 and 11 months in state 2. Ask 3

3. Life A is now 1 months in this state (***point to the appropriate colored card and say like… mention name of person on card***) followed by 11 months in perfect health (***point pink card and say like ….mention name of person on card***). What happens after is not known. Life B is immediate death (***like….mention name of person on card***). Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in one column or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1. Otherwise, ask 4***

Move slider to 10 months in state 1 and 2 months in state 2. Ask 4

4. Life A is 10 months in this state (***point to the appropriate colored card and say like… mention name of person on card***) followed by 2 months in perfect health (***point pink card and say like ….mention name of person on card).*** What happens after is not known. Life B is immediate death (***like….mention name of person on card***). Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in one column or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1. Otherwise, ask 5***

Move slider to 2 months in state 1 and 10 months in state 2. Ask 5

5. Life A is 2 months in this state (***point to the appropriate colored card and say like… mention name of person on card***) followed by 10 months in perfect health (***point pink card and say like ….mention name of person on card***). What happens after is not known. Life B is immediate death (***like….mention name of person on card***). Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in one column or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1. Otherwise, ask 6***

Move slider to 9 months in state 1 and 3 months in state 2. Ask 6

6. Life A is 9 months in this state (***point to the appropriate colored card and say like… mention name of person on card***) followed by 3 months in perfect health (***point pink card and say like ….mention name of person on card***). What happens after is not known. Life B is immediate death (***like….mention name of person on card***). Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in one column or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1. Otherwise, ask 7***

Move slider to 3 months in state 1 and 9 months in state 2. Ask 7

7. Life A is 3 months in this state (***point to the appropriate colored card and say like… mention name of person on card***) followed by 9 months in perfect health (***point pink card and say like ….mention name of person on card***). What happens after is not known. Life B is immediate death (***like….mention name of person on card***). Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in one column or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1. Otherwise, ask 8***

Move slider to 8 months in state 1 and 4 months in state 2. Ask 8

8. Life A is 8 months in this state (***point to the appropriate colored card and say like… mention name of person on card***) followed by 4 months in perfect health (***point pink card and say like ….mention name of person on card***). What happens after is not known. Life B is immediate death (***like….mention name of person on card***). Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in one column or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1. Otherwise, ask 9***

Move slider to 4 months in state 1 and 8 months in state 2. Ask 9

9. Life A is 4 months in this state (***point to the appropriate colored card and say like… mention name of person on card***) followed by 8 months in perfect health (***point pink card and say like ….mention name of person on card***). What happens after is not known. Life B is immediate death (***like….mention name of person on card***). Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in one column or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1. Otherwise, ask 10***

Move slider to 7 months in state 1 and 5 months in state 2. Ask 10

10. Life A is 7 months in this state (***point to the appropriate colored card and say like… mention name of person on card***) followed by 5 months in perfect health (***point pink card and say like ….mention name of person on card***). What happens after is not known. Life B is immediate death (***like….mention name of person on card***). Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in one column or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1. Otherwise, ask 11***

Move slider to 5 months in state 1 and 7 months in state 2. Ask 11

11. Life A is 5 months in this state (***point to the appropriate colored card and say like… mention name of person on card***) followed by 7 months in perfect health (***point pink card and say like ….mention name of person on card***). What happens after is not known. Life B is immediate death (***like….mention name of person on card***). Would you prefer A or B or are they equal? (***Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in one column or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1. Otherwise, ask 12***

Move slider to 6 months in state 1 and 6 months in state 2. Ask 12

12. Life A is 6 months in this state ***(point to the appropriate colored card and say like... mention name of person on card)*** followed by 6 months in perfect health ***(point pink card and say like ....mention name of person on card)***. What happens after is not known. Life B is immediate death ***(like....mention name of person on card)***. Would you prefer A or B or are they equal? ***(Tick under column A or B, or mark = in the middle column. If you have two consecutive ticks in one column or =, stop. Switch to script for states better than death. Remove the next scenario to be valued and start from 1. Otherwise, ask 13***

13. Tell me the amount of time in weeks you would be willing to live in this state to avoid immediate death. _____Weeks.

When you have finished valuing the six health states, say .....

THANK YOU. WE HAVE FINISHED WITH THAT METHOD

SCHISTOSOMIASIS HEALTH STATES VALUATION RECORD FORM

[This questionnaire must be completed in pencil]

1	Questionnaire number		

		Day		Month		Year			
2	Date of interview.								

3. Study number and phase ***[Tick the appropriate box]***

1	2	3
S1P1 (Pre-test)	S1P2A (Patient study)	S1P2B (Community study)

4.	Household number		

5.a Measurement study household number_____

5b. Measurement study serial number_____

6a Valuation study serial number (Initial study)_____

6b. Valuation study serial number (re-test study) _____

7. Name of interviewer_____

8.

Location *[Circle one only]*	1. Tebere
	2. Mutithi
	3. Thiba
	4. Nyangiti
	5. Murinduko

9. Sub-location *[Circle one only]*

1	Kiarukungu	12	Wamumu
2	Mahigaini	13	Thiba
3	Gathigiriri	14	Mathangauta
4	Rukanga	15	Kirimara
5	Kinyaga	16	Ndomba
6	Mathigaini	17	Nyangati
7	Kombuini	18	Riagicheru
8	Kathiga	19	Mugabaciura
9	Kiandegwa	20	Miuu
10	Kabiriri	21	Kamunyange
11	Nguka		

Record time interview starts (***fill the time in the appropriate row)***

				A.M
				P.M.

VISUAL ANALOG SCALE (VAS) RECORD FORM

The interviewer should answer questions 9 and 10 after going through the VAS example

10. How many times did you explain the VAS procedure to the respondent? _____Times

11. Record time VAS exercise starts (***fill the time in the appropriate row)***

				A.M
				P.M.

VAS HEALTH STATES RANKING SCORE SHEET

(Use this table to fill in values of the health state scenarios as the valuation exercise proceeds. Allow the respondent to change their rankings as desired. Columns 3 and 4 of this table should be completed at the end of the day while checking for completeness of the questionnaires)

HEALTH STATE	VAS FEELING THERMOMETER SCORE	STATE BETTER OR WORSE THAN DEATH *(Fill in the appropriate case against state)*	INCONSISTENT RANKING? *(indicate yes or no against appropriate state)*
PH			
A			
C			
E			
F			
ZZ			

Complete the visual analog scale ranking scoring sheet on the next page by transferring the health states as the respondent arranged them on the large portable VAS FEELING THERMOMETER. Make sure the arrows from the health states to feeling thermometer clearly point the respondent's value when transferred to the VAS on the record form).

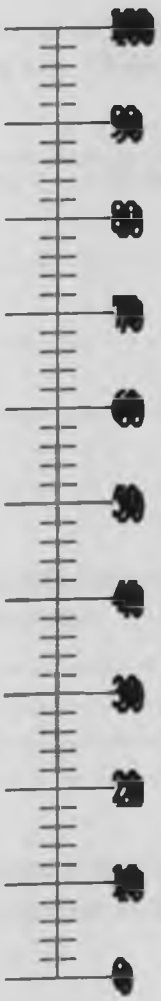
12. Record time VAS exercise is finished (***fill the time in the appropriate row)***

				A.M
				P.M.

VISUAL ANALOG SCALE RANKING SCORING SHEET

Household No. / patient No.

Best imaginable health state



Worst imaginable health state

PH

A

C

E

F

ZZ

In the next three questions, I want to find out what you feel about the method we have just used.

13. Did you find the VAS method (***Circle one only***)
    Very easy to understand
    Easy to understand
    Fairly easy to understand
    Difficult to understand
    Very difficult to understand?

14. Tell me all the kinds of people who might object the use of the VAS method in your community? (***For each kind of person ask, "Why would this kind of person object the use of VAS?" record verbatim***)

Kind of person who might object	***Reason they might object***

15. Tell me in your own words the difficulties you experienced in going through the VAS exercise we have just completed.

_____
_____
_____
_____

THIS SECTION SHOULD BE COMPLETED BY THE INTERVIWER AT THE END OF VAS EXERCISE. (*Your responses to this section should be based on your impression of the interview).*

16. How would you rate the respondent's level of difficult in understanding the VAS exercise? (***Circle one only***)
    No difficult at all
    A little difficult
    Somewhat difficult
    Difficult
    Very difficult

17. How comfortable was the respondent with the way the VAS questions were posed to him/her?

    Not comfortable at all
    A little comfortable
    Somewhat comfortable
    Comfortable
    Very comfortable.

18. What is your impression about the interview and the responses that you got?

    Very good
    Good
    Satisfactory
    Poor
    Very poor.

TIME TRADE OFF RECORD FORM

Record time TTO process starts (***fill the time in the appropriate row)***

				A.M
				P.M.

(***Interviewer should complete 19-22 after explaining and establishing that the respondent has understood the procedure)***

19. Does the respondent understand the TTO procedure for states better than death?
    1. Yes (***Do valuation interview)*** 2. No (***Do not interview)***

20. Does the respondent understand the TTO procedure for states worse than death?
    1. Yes (***Do valuation interview)*** 2. No (***Do not interview)***

21. How many times did you explain the TTO procedure for states better than death to the respondent? _____ Times.

22. How many times did you explain the TTO procedure for states worse than death to the respondent? _____ Times.

TTO INTERVIEW

23. Record time actual TTO valuation exercise starts (***fill the time in the appropriate row)***

				A.M
				P.M.

Record the respondent's choice by indicating with a tick either choice A or choice B, or whether they are equal in the appropriate columns. If you get two consecutive ticks in the first two rows of column A sheet 1, switch to script for states worse than death and use sheet 2. Otherwise, continue the interview up to the point where you get two consecutive ticks in either column A or B or where the respondent indicates the two choices are equal. Then remove the next health state scenario to be valued and repeat the procedure.

24. HEALTH STATE SCENARIO A [ANTHONY / ANNE]

(Use sheet 1 if state considered better than death. Use sheet 2 if state considered worse than death)

SHEET 1

			Life A or life B *(Tick which option the respondent prefers)*		
	Time in life A	Time in life B	A	A and B equal	B
Either live in life A for _____ months. OR Live in life B for _____ months. Which do you prefer?	12	12			
	0	12			
	11	12			
	1	12			
	10	12			
	2	12			
	9	12			
	3	12			
	8	12			
	4	12			
	7	12			
	5	12			
	6	12			

SHEET 2

			Life A or life B *(Tick which option the respondent prefers)*		
	Time in life A	Time in life B	A	A and B equal	B
Either live in life A for _____ months in this state and then _____months in perfect health. OR Die immediately. Which do you prefer?	0 then 12	0			
	11 then 1	0			
	1 then 11	0			
	10 then 2	0			
	2 then 10	0			
	9 then 3	0			
	3 then 9	0			
	8 then 4	0			
	4 then 8	0			
	7 then 5	0			
	5 then 7	0			
	6 then 6	0			

(Interviewer to answer questions 1-5 after completing valuing each scenario. Questions 2 and 3 should only be asked if Q14 in the better than death and question 13 in the worse than death interview scripts are asked. Ask the respondent to answer question 6)

1. State considered ***(circle one).***
   1.Better than dead ***(ask 2 if you asked 14).*** 2.Worse than dead ***(ask 3 if you asked 13)***

2. What amount of time in weeks are you willing to give up in perfect health to avoid living in life B? _____Weeks.

3. What amount of time in weeks are you willing to live in this state to avoid immediate death? _____Weeks.

4. Health state Value. _____ ***(Do not fill in. To be computed when checking for completeness of questionnaires)***

5. Did the respondent choose worse off health state scenario over a better of health state scenario state? (*For example choosing any of the states being valued rather than perfect health OR preferring death to any of the states being valued)*

   1 Yes ***(Ask 6)*** 2 No (***Go to next scenario to be valued)***

6. Please tell me your reasons for choosing this health state (*mention the identifying letter of the worse of state)* over this one *(Mention the identifying letter of the better off health state).*

Chose this instead of this		***Reason for choosing a worse-off state instead of a better-off state***

25. HEALTH STATE SCENARIO F [FRANCIS / FAITH]

(Use sheet 1 if state considered better than death. Use sheet 2 if state considered worse than death)

SHEET 1

			Life A or life B *(Tick which option the respondent prefers)*		
	Time in life A	Time in life B	A	A and B equal	B
Either live in life A for _____ months. OR Live in life B for _____ months. Which do you prefer?	12	12			
	0	12			
	11	12			
	1	12			
	10	12			
	2	12			
	9	12			
	3	12			
	8	12			
	4	12			
	7	12			
	5	12			
	6	12			

SHEET 2

			Life A or life B *(Tick which option the respondent prefers)*		
	Time in life A	Time in life B	A	A and B equal	B
Either live in life A for _____ months in this state and then _____months in perfect health. OR Die immediately. Which do you prefer?	0 then 12	0			
	11 then 1	0			
	1 then 11	0			
	10 then 2	0			
	2 then 10	0			
	9 then 3	0			
	3 then 9	0			
	8 then 4	0			
	4 then 8	0			
	7 then 5	0			
	5 then 7	0			
	6 then 6	0			

(Interviewer to answer questions 1-5 after completing valuing each scenario. Questions 2 and 3 should only be asked if Q14 in the better than death and question 13 in the worse than death interview scripts are asked. Ask the respondent to answer question 6)

1. State considered ***(circle one).***
   1.Better than dead ***(ask 2 if you asked 14).*** 2.Worse than dead ***(ask 3 if you asked 13)***

2. What amount of time in weeks are you willing to give up in perfect health to avoid living in life B? _____Weeks.

3. What amount of time in weeks are you willing to live in this state to avoid immediate death? _____Weeks.

4. Health state Value. _____ ***(Do not fill in. To be computed when checking for completeness of questionnaires)***

5. Did the respondent choose worse off health state scenario over a better of health state scenario state? (*For example choosing any of the states being valued rather than perfect health OR preferring death to any of the states being valued)*

   1 Yes ***(Ask 6)*** 2 No (***Go to next scenario to be valued)***

6. Please tell me your reasons for choosing this health state (*mention the identifying letter of the worse of state)* over this one *(Mention the identifying letter of the better off health state).*

Chose this instead of this		***Reason for choosing a worse-off state instead of a better-off state***

26. HEALTH STATE SCENARIO C [CHRISTOPHER / CAROL]

(Use sheet 1 if state considered better than death. Use sheet 2 if state considered worse than death)

SHEET 1

			Life A or life B *(Tick which option the respondent prefers)*		
	Time in life A	Time in life B	A	A and B equal	B
Either live in life A for _____ months. OR Live in life B for _____ months. Which do you prefer?	12	12			
	0	12			
	11	12			
	1	12			
	10	12			
	2	12			
	9	12			
	3	12			
	8	12			
	4	12			
	7	12			
	5	12			
	6	12			

SHEET 2

			Life A or life B *(Tick which option the respondent prefers)*		
	Time in life A	Time in life B	A	A and B equal	B
Either live in life A for _____ months in this state and then _____months in perfect health. OR Die immediately. Which do you prefer?	0 then 12	0			
	11 then 1	0			
	1 then 11	0			
	10 then 2	0			
	2 then 10	0			
	9 then 3	0			
	3 then 9	0			
	8 then 4	0			
	4 then 8	0			
	7 then 5	0			
	5 then 7	0			
	6 then 6	0			

(Interviewer to answer questions 1-5 after completing valuing each scenario. Questions 2 and 3 should only be asked if Q14 in the better than death and question 13 in the worse than death interview scripts are asked. Ask the respondent to answer question 6)

1. State considered ***(circle one).***
   1.Better than dead ***(ask 2 if you asked 14).*** 2.Worse than dead ***(ask 3 if you asked 13)***

2. What amount of time in weeks are you willing to give up in perfect health to avoid living in life B? _____Weeks.

3. What amount of time in weeks are you willing to live in this state to avoid immediate death? _____Weeks.

4. Health state Value. _____ ***(Do not fill in. To be computed when checking for completeness of questionnaires)***

5. Did the respondent choose worse off health state scenario over a better of health state scenario state? (*For example choosing any of the states being valued rather than perfect health OR preferring death to any of the states being valued)*

   1 Yes ***(Ask 6)*** 2 No (***Go to next scenario to be valued)***

6. Please tell me your reasons for choosing this health state (*mention the identifying letter of the worse of state)* over this one *(Mention the identifying letter of the better off health state).*

Chose this instead of this		***Reason for choosing a worse-off state instead of a better-off state***

27. HEALTH STATE SCENARIO E [ERIC / EMMA]

(Use sheet 1 if state considered better than death. Use sheet 2 if state considered worse than death)

SHEET 1

			Life A or life B *(Tick which option the respondent prefers)*		
	Time in life A	Time in life B	A	A and B equal	B
Either live in life A for _____ months. OR Live in life B for _____ months. Which do you prefer?	12	12			
	0	12			
	11	12			
	1	12			
	10	12			
	2	12			
	9	12			
	3	12			
	8	12			
	4	12			
	7	12			
	5	12			
	6	12			

SHEET 2

			Life A or life B *(Tick which option the respondent prefers)*		
	Time in life A	Time in life B	A	A and B equal	B
Either live in life A for _____ months in this state and then _____months in perfect health. OR Die immediately. Which do you prefer?	0 then 12	0			
	11 then 1	0			
	1 then 11	0			
	10 then 2	0			
	2 then 10	0			
	9 then 3	0			
	3 then 9	0			
	8 then 4	0			
	4 then 8	0			
	7 then 5	0			
	5 then 7	0			
	6 then 6	0			

(Interviewer to answer questions 1-5 after completing valuing each scenario. Questions 2 and 3 should only be asked if Q14 in the better than death and question 13 in the worse than death interview scripts are asked. Ask the respondent to answer question 6)

1. State considered ***(circle one).***
   1.Better than dead ***(ask 2 if you asked 14).*** 2.Worse than dead ***(ask 3 if you asked 13)***

2. What amount of time in weeks are you willing to give up in perfect health to avoid living in life B? _____Weeks.

3. What amount of time in weeks are you willing to live in this state to avoid immediate death? _____Weeks.

4. Health state Value. _____ ***(Do not fill in. To be computed when checking for completeness of questionnaires)***

5. Did the respondent choose worse off health state scenario over a better of health state scenario state? (*For example choosing any of the states being valued rather than perfect health OR preferring death to any of the states being valued)*

   1 Yes ***(Ask 6)*** 2 No (***Go to next scenario to be valued)***

6. Please tell me your reasons for choosing this health state (*mention the identifying letter of the worse of state)* over this one *(Mention the identifying letter of the better off health state).*

Chose this instead of this		***Reason for choosing a worse-off state instead of a better-off state***

28. Record time TTO exercise ends (***fill the time in the appropriate row)***

				A.M
				P.M.

29. Did you find the TTO method (***Circle one only)***

1 Very easy to understand
2 Easy to understand
3 Fairly easy to understand
4 Difficult to understand
5. Very difficult to understand?

30. Tell me all the kinds of people who might object the use of the TTO method in your community? (***For each kind of person ask,*** "Why would this kind of person object the use of TTO?" ***record verbatim)***

Kind of person who might object	***Reason they might object***

31. Tell me in your own words the difficulties you experienced in going through the TTO exercise we have just completed.

_____
_____
_____
_____

THIS SECTION SHOULD BE COMPLETED BY THE INTERVIWER AT THE END OF TTO EXERCISE. (*Your responses to this section should be based on your impression of the interview).*

32. How would you rate the respondent's level of difficult in understanding the TTO exercise? (***Circle one only)***

1. No difficult at all
2. A little difficult
3. Somewhat difficult
4. Difficult
5. Very difficult

33. How comfortable was the respondent with the way the TTO questions were posed to him/her?

1. Not comfortable at all
2. A little comfortable
3. Somewhat comfortable
4. Comfortable
5. Very comfortable

1. What is your impression about the TTO interview and the responses that you got?
2. Very good
3. Good
4. Satisfactory
5. Poor
6. Very poor

GENARAL EVALUATIVE QUESTIONS

35. How did you find each of the two methods that we have used to value these health states, in terms of ease of understanding the questions and the exercise in general. ***(Tick one box against each method)***

	1=Easy	2=Neither easy nor difficult	3=Difficult
Visual Analog Scale			
Time trade off			

36. If we were to repeat this valuation exercise, using only one of the two methods, which method would you prefer to use? ***(Circle one to indicate the most preferred)***

1	**Visual Analog Scale**
2	**Time trade off**

37. Please tell me what led you to make your choice above. (***Write a separate explanation each method).***

1._____

_____

_____

_____

2._____

_____

_____

_____

38. For each of the method that we have used, please indicate whether you thought

1= very hard.
2=hard.
3= A little.
4= very little.

when I asked you the questions.

TTO_____
VAS_____

PERSONAL DATA RECORD FORM

To help in comparing responses from the various persons we have talked to I would like to ask you some general questions about yourself in this last section.

39 Name *(optional)*_____

40. Sex: 1. Male 2. Female *[circle one]*

41. What is your age? .....*(years).*

42. Do you have any children? 1. Yes. *(Ask 63)* 2. No. *(Ask 64)*

43. If yes, how many? _____Child(ren).

44. What is your current marital status? *[circle the correct response]*

1. Single 2. Married 3. Co-habiting
4. Divorced /Separated 5. Widower/ Widow 6. Other (specify)

What is the highest level of education that you have reached? *[Circle only one]*

	Educational level attained
1	None *[Do not ask 135]*
2	Primary
3	Secondary ("O" level)
4	Secondary ("A" level)
5	Diploma
6	Graduate (degree)
7	Adult education
8	Other (specify)

46.

How many years did you spend in *[Indicate number of years spent in school against each level up to the highest level reached as reported in 134 above]*	**Educational level attained**	**No. of years**
	Primary?	
	Secondary ("O" level)?	
	Secondary ("A" level)?	
	Diploma?	
	Graduate (degree)?	
	Adult education?	
	Other (specify)?	
Total years in school *[Do not add during interview. Add up after]*		

47. What is your main profession?

1 Farmer
2 Teacher
3 Businessperson (specify type)
4 Casual laborer
5 Civil servant (specify)
6 Student
7 other (specify)

48. What do you do for a living?

1 Farmer
2 Teacher
3 Businessperson (Specify type)
4 Casual laborer
5 Civil servant (specify)
6 Student
7. Other (specify)

49. Please tell me how much your household spends per month on each of the following expenditure items. (***For those respondents under working age and or single, ask this question if they indicate they know the information. If a respondent casts doubt over their knowledge of this information do not press for an answer. Note besides the question the reaction of the respondent when you posse this question).***

code	Expenditure item	Amount spent per month (Kshs)
1	Food	
2	Clothing	
3	Rent	
4	School fees ***[divide term fees by 3]***	
5	Health	
6	Debt repayment	
7	Bills (water, electricity, telephone)	
8	Leisure	
9	Others (specify)	

50. I will read out for you different groups of levels of income. Please tell me the level of monthly income that best describes your household's total earnings ***(do not ask those under working age. e.g those still in school and hence dependants, except if they report they are working or they know the information).***

Code	Income level (Kshs.)	(Tick)
1	0-2,000	
2	2,001-5,000	
3	5,001-10,000	
4	10,001-20,000	
5	20,001-50,000	
6	Above 50,000	

51. Current health state (***Please tell me the statement that best describes how you have been feeling about your health state in the last two weeks***)

Circle only one statement in each domain that describes how the respondent has feeling about his/her health state in the last two weeks.

1. In the last two weeks have you been unable to walkabout, go to workplace and move about without problems;
    1. None of the time
    2. A little of the time
    3. Some of the time
    4. Most of the time
    5. All the time

2. In the last two weeks have you been unable to perform your work and daily duties (e.g. playing, farming, household chores, usual activity [activity done for livelihood] schoolwork, etc);
    1. None of the time
    2. A little of the time
    3. Some of the time
    4. Most of the time
    5. All the time

3. In the last two weeks has your performance and output of work i.e. being able to accomplish as much as desired in activities of livelihood) been affected;
    1. None of the time
    2. A little of the time
    3. Some of the time
    4. Most of the time
    5. All the time

4. In the last two weeks you have not felt strong and energetic
    1. None of the time
    2. A little of the time
    3. Some of the time
    4. Most of the time
    5. All the time

5. In the last two weeks have you been unable to participate in social functions such as group meetings, religious meetings, weddings, visiting friends and family and socializing with friends and colleagues;
    1. None of the time
    2. A little of the time
    3. Some of the time
    4. Most of the time
    5. All the time

6. In the last two weeks you felt worry and anxiety;
   1. None of the time
   2. A little of the time
   3. Some of the time
   4. Most of the time
   5. All the time

Health state description (To be completed at the end of the interview)

52. Tell me the number of days you experienced any of the following symptoms in the last two weeks. ***(Read out each symptom for the respondent and tick only one frequency against each symptom)***

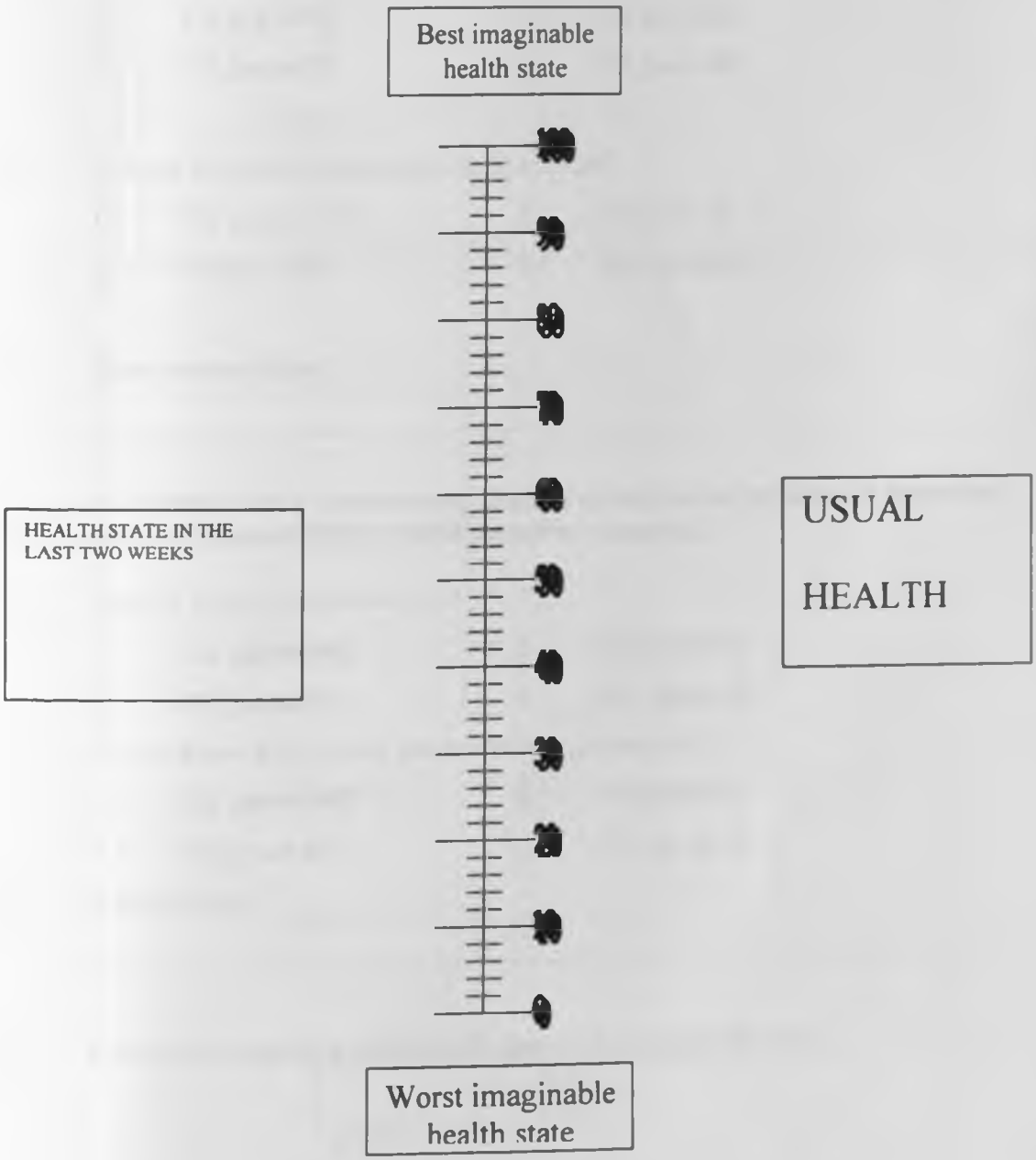
Symptoms	**NA**	**frequency**				
		1=less than a day	**2=1-3 days**	**3=4-6 days**	**4=7-10 days**	**5=1-14 days**
Abdominal pain and discomfort						
Watery diarrhea						
Bloody mucoid diarrhea						
Tiredness						
Loss of appetite						
Itching skin rash						

53. For each of the symptoms you experienced in the last two weeks, tell me the term that best describes its intensity at its worst. (***Read out each term against each symptom mentioned above and circle only one intensity description)***

Symptoms	**NA**	**Intensity**			
Abdominal pain and discomfort		mild	moderate	severe	Very severe
Watery diarrhea *(No. of times stool passed per day)*		3-9 times a day	10-15 times a day	16-21 times a day	Over 21 times a day
Bloody mucoid diarrhea *(No. of times stool passed per day)*		3-9 times a day	10-15 times a day	16-21 times a day	Over 21 times a day
Tiredness		A little tired	Somewhat tired	Very tired	Extremely tired
Loss of appetite (*Proportion of the amount normally eaten one was able to eat.)*		About three quarters	Half to a quarter the amount	Not more than two spoonfuls	None at all
Itching skin rash		Mild	Moderate	severe	Very severe

54. Where on this scale would you rate your health state in the last two weeks and your usual health state.

VISUAL ANALOG SCALE RANKING SCORING SHEET



HISTORY OF CURRENT ILLNESS

The next 3 questions are related to your recent history of illness. In answering these questions think of the last 2 weeks ***[Circle the correct response.]***

55. In the last two weeks, did you experience any illness?

1. Yes ***[go to56].*** 2. No. ***[go to58]***
3. DN ***[go to58]*** 4. NR ***[go to58]***

56. Did you know what illness you suffered from?

1. Yes ***[go to 57].*** 2. No ***[go to58]***
3. DN ***[go to58]*** 4. NR ***[go to58]***

57. What was the illness?

_____

The next 3 questions are related to any illness you may be having now. In answering these questions think of <u>NOW</u> ***[Circle the correct response.]***

58. Are you having any illness now?

1. Yes ***[go to 59].*** 2. No ***[go to 61].***
3. DN ***[go to 61].*** 4. NR ***[go to 61].***

59 Do you know what illness you are suffering from now?

1. Yes ***[go to 60].*** 2. No ***[go to 61].***
3. DN ***[go to 61].*** 4. NR ***[go to 61].***

60 Which illness?

_____

61. Record time interview ends ***(fill the time in the appropriate row)***

				A.M
				P.M.

REQUEST TO PARTICIPATE IN THE NEXT ROUND OF THE STUDY

Thank you for answering all my questions. Before we finish, I would like to make a request from you. To enable us know whether these methods that we have used today can be used any other time and give the same values for the health states we have valued, we would like to repeat this exercise again in the coming four or so weeks. We would very much like you to participate in that exercise as a follow-up of this one. I am now kindly requesting you to tell me whether you are willing to be contacted for the next round of this study.

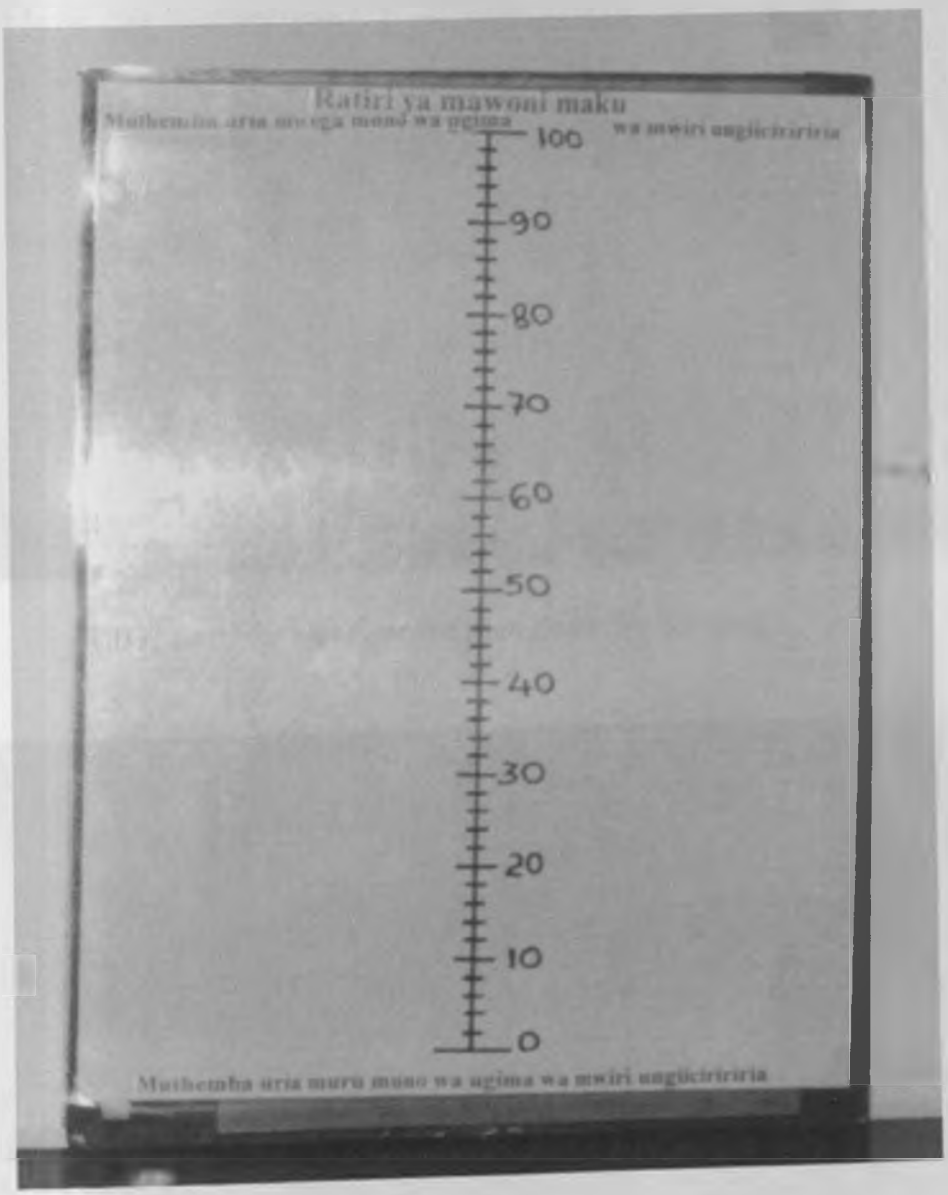
62. Would you be willing to participate in the next round of this study?

Yes *[Thank you and end]*	No *[Thank you and end]*	DN *[Thank you and end]*	NR *[Thank you and end]*

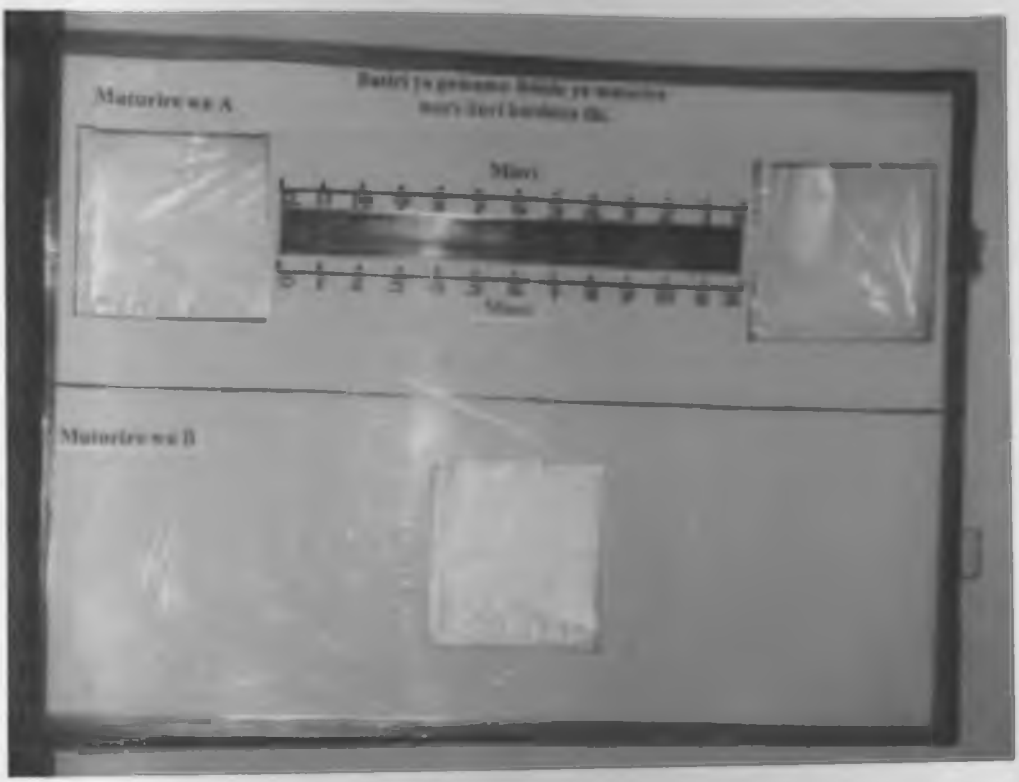
We have now come to the end of this interview. I thank you very much for all you help and time. Is there anything you would like to ask me before we finish?

THANK YOU ONCE AGAIN FOR YOUR TIME AND ASSISTANCE.

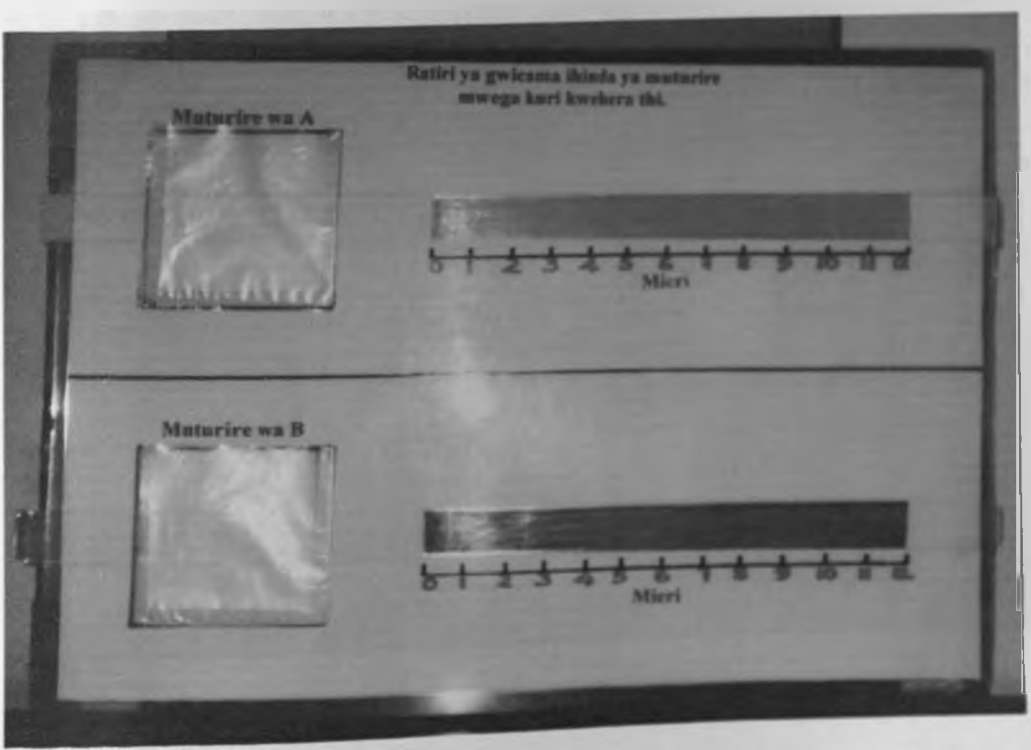
APPENDIX 8.2: PROPS: VAS AND TTO BOARDS



VAS Feeling thermometer (in Kikuyu)



TTO board for states worse than death (in Kikuyu)



TTO board for states better than death (in Kikuyu)

APPENDIX 8.3: COLOUR CODED HEALTH STATES

PH Perfect Health PAUL /PERIS

- He/she has no symptoms at all.
- His/ her performance of work and daily duties and social participation are affected none of the time. He/she feels worry and anxiety none of the time.

A ANTHONY /ANNE

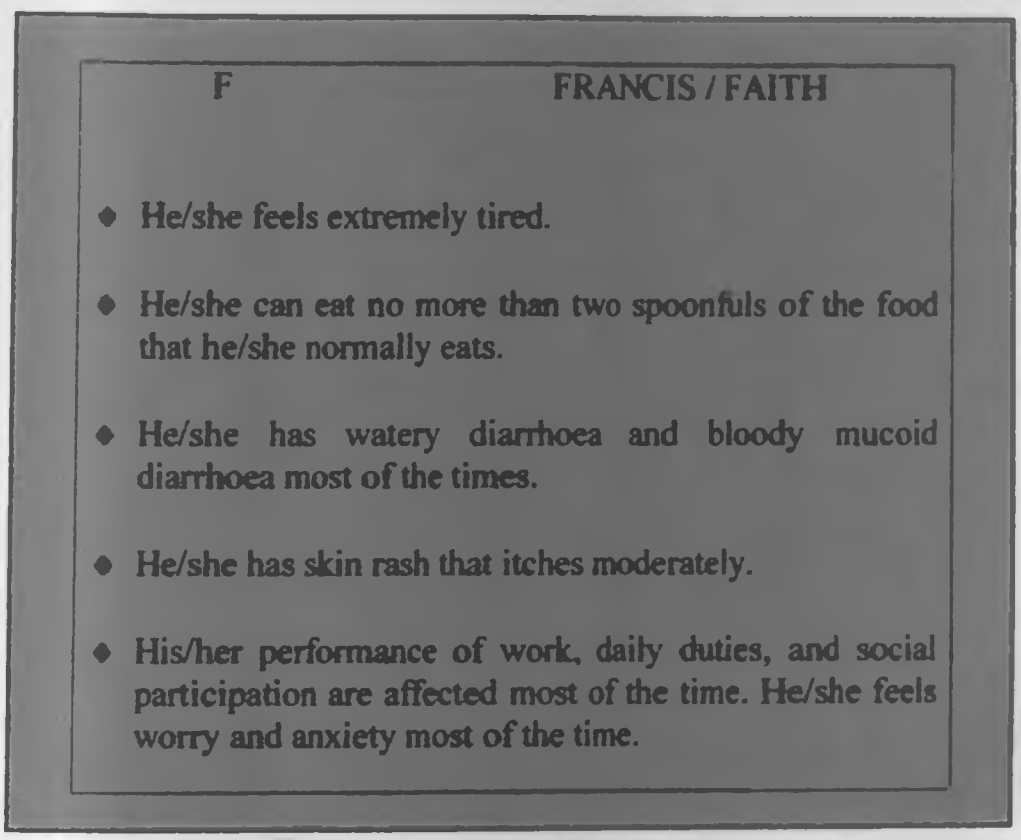
- He/she feels somewhat tired
- He/she can only eat about half to a quarter the amount of food that he/she normally eats.
- His/her performance of work, daily duties, and social participation are affected a little of the time. He/she feels worry and anxiety a little of the time.

C CHRISTOPHER / CAROL

- He/she feels somewhat tired and then very tired.
- He/she can only eat half to a quarter of the food he/she you normally eats and at times no more than two spoonfuls the amount of food that he/she normally eats.
- His/her performance of work, daily duties and social participation are affected some of the time. He/she feels worry and anxiety some of the time.

E ERIC / EMMA

- He/she feels very tired.
- He/she can only eat half to a quarter of the food that he/she normally eats and at times no more than two spoonfuls of the food that he/she normally eats.
- He/she has watery diarrhoea and bloody mucoid diarrhoea sometimes.
- He/she has skin rash that itches moderately.
- His/ her performance of work, daily duties, and social participation are affected most of the time. He/she feels worry and anxiety most of the time.


F FRANCIS / FAITH
♦ He/she feels extremely tired.
♦ He/she can eat no more than two spoonfuls of the food that he/she normally eats.
♦ He/she has watery diarrhoea and bloody mucoid diarrhoea most of the times.
♦ He/she has skin rash that itches moderately.
♦ His/her performance of work, daily duties, and social participation are affected most of the time. He/she feels worry and anxiety most of the time.


ZZ ZACHARIA / ZIPPORAH
♦ DEAD

MM MOCK HEALTH STATE

- He/she feels somewhat tired and at times very tired
- He/she can eat not more than two spoonfuls of the food that he/she normally eats
- He/she has watery diarrhoea and bloody mucoid diarrhoea most of the time
- He/she has skin rash that itches moderately
- His/her performance of output and work, and social participation are affected some of the time. He/she feels worry and anxiety some of the time.

APPENDIX 8.4: TABLES AND CHARTS

Figure A8.1: Distribution of HRQL states amongst study

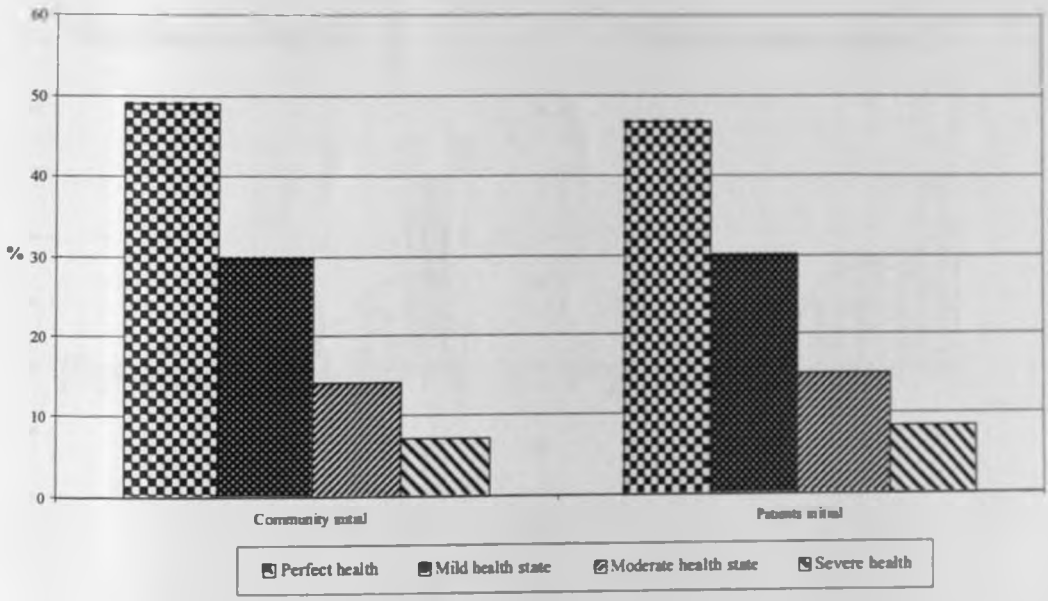


Figure A8.2: Comparisons of distributions of VAS and TTO values for *S. Mansoni* disease states (n=117)

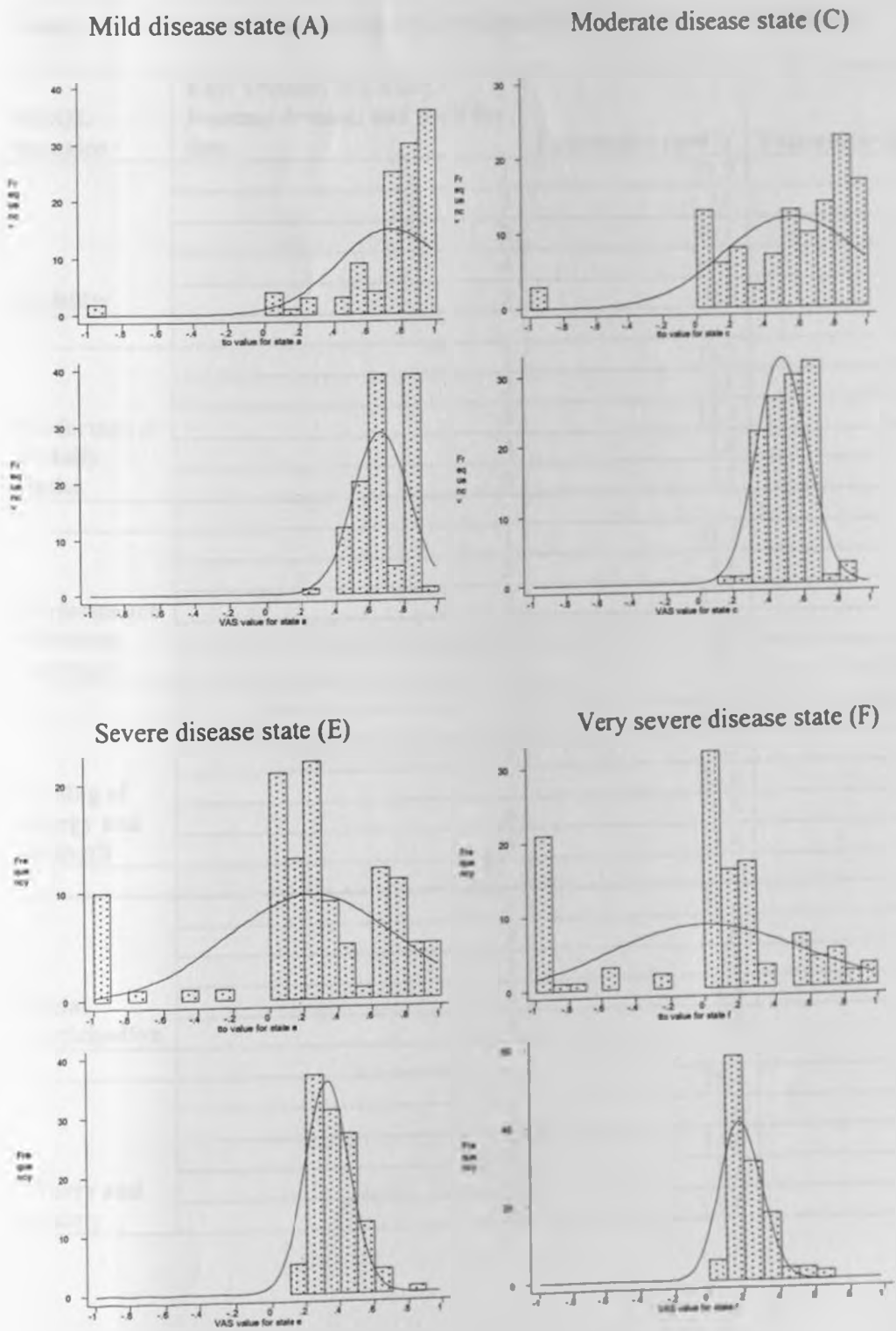


Table A8.1: Current health state-: How often HRQL domains were affected

HRQL domains	Key: 1=none; 2=a little; 3=some; 4=most; and 5=all the time	Community (n=57)	Patients (n=60)
Mobility	1	71.9	73.3
	2	7.0	3.3
	3	10.5	18.3
	4	10.5	5.0
	5	0.0	0.0
Performance of daily duties	1	73.7	68.3
	2	8.8	3.3
	3	12.3	20.0
	4	1.8	6.7
	5	3.5	1.7
Performance of output and work	1	63.2	71.7
	2	8.8	6.7
	3	21.1	15.0
	4	3.5	6.7
	5	3.5	0.0
Feeling of energy and strength	1	64.9	53.3
	2	5.3	8.3
	3	22.8	23.3
	4	5.3	8.3
	5	1.8	6.7
Social participation	1	84.2	83.3
	2	7.0	3.3
	3	3.5	8.3
	4	1.8	5.0
	5	3.5	0.0
Worry and anxiety	1	75.4	75.0
	2	7.0	5.0
	3	14.0	15.0
	4	3.5	5.0
	5	0.0	0.0

Table A8.2: Aspects of practicality of VAS and TTO

		Community(n=57)		Patients(n=60)	
		VAS	TTO	VAS	TTO
Extent of thinking required to use technique					
	Very hard	7.0	8.8	8.3	8.3
	Hard	21.1	50.9	33.3	46.7
	A little	63.2	28.1	50.0	40.0
	Very little	8.8	12.3	8.3	5.0
Time to complete valuation task (minutes): mean (SD) [range]		9.0 (3.8) [4-23]	16.9 (8.0) [5-45]	6.9 (3.1) [3-18]	13.0 (4.5) [4-27]
Groups of people who cannot use technique (% holding opinion)					
	Young children	91.2	91.2	100.0	98.3
	Very old	75.4	77.2	95.0	91.7
	Illiterate	43.9	35.1	28.3	26.7
	Mentally unstable	19.3	19.3	20.0	13.3
	Drug abusers	8.77	0.0	0.0	0.0
	Sick	22.8	22.8	38.3	20.0
Age of children and old who cannot use technique: mean (SD) [range]					
	Children	9.5 (4.4) [5-18]	9.3 (4.7) [5-18]	9.8 (3.3) [5-25]	9.9 (4.2) [5-25]
	Old	52.1 (33.2) [45-100]	50.7 (35.2) [40-100]	66.5 (23.0) [40-100]	62.8 (25.2) [40-100]

Table A8.3: Mean VAS and TTO values at initial valuation (n=117)

STATES		VAS	TTO
Perfect health	Mean	0.96	n.a
	Standard deviation	0.08	n.a
	Range	0.58-1	n.a
	95% C.I.	0.94-0.97	n.a
State A	Mean	0.67	0.73
	Standard deviation	0.16	0.31
	Range	0.2-0.92	-0.96 - 1
	95% C.I.	0.63-0.69	0.68-0.79
State C	Mean	0.48	0.55
	Standard deviation	0.14	0.39
	Range	0.14-0.82	-0.96 - 1
	95% C.I.	0.46-0.51	0.48-0.62
State E	Mean	0.32	0.24
	Standard deviation	0.13	0.48
	Range	0.1-0.8	-0.96 - 0.96
	95% C.I.	0.3-0.35	0.15-0.33
State F	Mean	0.18	0
	Standard deviation	0.12	0.54
	Range	0-0.6	-0.96 - 0.96
	95% C.I.	0.16-0.20	-0.1 - 0.09
Dead	Mean	0	n.a
	Standard deviation	0.06	n.a
	Range	0 -0.7	n.a
	95% C.I.	0-0.02	n.a

n.a= not applicable

Table A8.4: Inconsistencies in ranking of disease states (n=117)

		VAS	TTO
No. of Disease states ranked consistently with disease severity	0		11.1
	1	1.7	16.2
	2	0.0	29.1
	3	0.9	21.4
	4	2.6	22.2
	5	2.6	n.a
	6	92.3	n.a
	Mean (SD) [range]	6 (0.8) [1-6]	2.3 (1.3) [0-4]
% ranking disease state inconsistently			
	State PH	0.9	n.a -
	State A	2.6	1.7
	State C	4.3	2.6
	State E	4.3	11.1
	State F	6.0	23.9
	State ZZ	1.7	n.a -

Key: PH-perfect health; A-Mild disease state; C-moderate disease state; E-severe disease state; F-very severe disease state, ZZ- dead
n.a = not applicable

Table A8.5: Differences in mean values between males and females (n=117)

	VAS			TTO		
STATES	male	Female	Mean difference	male	Female	Mean difference
State A	0.61	0.70	-0.09[ab]	0.78	0.71	0.07
State C	0.46	0.49	-0.04	0.63	0.50	0.12[a]
State E	0.32	0.32	0.00	0.28	0.22	0.06
State F	0.20	0.16	0.04[a]	0.04	-0.03	0.07

a- p<0.05 using t test. b – p<0.05 using Mann-Whitney test

Table A8.6: Differences in mean values between age groups (n=117)

	VAS mean values for disease states			
Age in yrs	State A	State C	State E	State F
15-30	0.67	0.48	0.33	0.19
31-45	0.67	0.46	0.31	0.16
46-60	0.64	0.47	0.31	0.18
61-77	0.67	0.57	0.34	0.18
	TTO mean values for disease states			
	State A	State C	State E	State F
15-30	0.77	0.59	0.22	-0.04
31-45	0.66	0.56	0.28	0.08
46-60	0.81	0.55	0.25	-0.03
61-77	0.62	0.41	0.25	0.06

No significant differences using Kruskal Wallis test of one-way ANOVA F statistic

Table A8.7: Differences in mean values between marital status groups (n=117)

	VAS mean values for disease states			
Marital status	**State A**	**State C**	**State E**	**State F**
Single	0.67	0.47	0.33	0.18
Married	0.66	0.48	0.33	0.18
Divorced/separated	0.62	0.43	0.30	0 13
Widow/widower	0.72	0.56	0.27	0.14
	TTO mean values for disease states			
	State A	**State C**	**State E**	**State F**
Single	0.76	0.58	0.18	-0.03
Married	0.75	0.55	0.28	-0.01
Divorced /separated	0.65	0.66	0.04	0.12
Widow/ widower	0.62	0.48	0.26	0.11

No significant differences using Kruskal Wallis test of one-way ANOVA F statistic

Table A8.8: Differences in mean values between those with children and not (n=117)

	VAS			TTO		
STATES	**with**	**without**	**Mean difference**	**with**	**without**	**Mean difference**
State A	0.67	0.65	0.02	0.73	0.78	-0.05
State C	0.48	0.47	0.01	0.53	0.63	-0.10
State E	0.32	0.33	-0.01	0.26	0.20	0.06
State F	0.18	0.19	-0.01	0.00	-0.01	0.01

No significant differences using t test and Mann-Whitney test

Table A8.9: Differences in mean values between grouped number of children (n=90)

	VAS mean values for disease states			
Number of children	**State A**	**State C**	**State E**	**State F**
1-3	0.69	0.49	0.34	0.19
4-6	0.65	0.47	0.28	0.16
7-9	0.63	0.49	0.32	0.16
10-15	0.66	0.50	0.28	0.20
	TTO mean values for disease states			
	State A [a]	**State C** [a]	**State E**	**State F**
1-3	0.77	0.60	0.27	-0.06
4-6	0.53	0.40	0.18	-0.06
7-9	0.80	0.38	0.28	0.12
10-15	0.89	0.82	0.30	0.27

a- $p<0.05$ using one-way ANOVA F statistic. No significant difference using Kruskal Wallis test.

Table A8.10: Differences in mean values between levels of education (n=117)

	VAS mean values for disease states			
Education level	**State A** [ab]	**State C** [b]	**State E**	**State F**
None	0.74	0.59	0.37	0.17
Primary	0.68	0.48	0.33	0.19
Secondary	0.59	0.42	0.29	0.16
Post secondary	0.68	0.50	0.25	0.10
	TTO mean values for disease states			
	State A [a]	**State C**	**State E**	**State F**
None	0.49	0.40	0.15	-0.1
Primary	0.77	0.53	0.19	0.04
Secondary	0.82	0.69	0.45	-0.08
Post secondary	0.58	0.83	0.27	0.13

a- $p<0.05$ using one-way ANOVA F statistic. b – $p<0.05$ using Kruskal Wallis test

Table A8.11: Differences in mean values between levels of monthly expenditure (n=117)

	VAS mean values for disease states			
Monthly household expenditure (KShs.)	**State A**	**State C**	**State E**	**State F**
0-2000	0.67	0.48	0.29	0.16
2001-5000	0.68	0.54	0.40	0.24
5001-10000	0.67	0.47	0.31	0.17
10001-20000	0.66	0.47	0.33	0.19
Above 20,000	0.64	0.50	0.32	0.17
	TTO mean values for disease states			
	State A	**State C**	**State E**	**State F**
0-2000	0.79	0.58	0.22	0.05
2001-5000	0.72	0.52	0.31	0.16
5001-10000	0.72	0.59	0.17	-0.08
10001-20000	0.72	0.49	0.31	-0.05
Above 20,000	0.79	0.54	0.35	0.15

No significant differences using Kruskal Wallis test of one-way ANOVA F statistic

Table A8.12: Differences in mean values between levels of monthly income (n=97)

	VAS mean values for disease states			
Monthly household income (KShs.)	**State A**	**State C**	**State E**	**State F**
0-2000	0.71	0.47	0.33	0.18
2001-5000	0.66	0.49	0.32	0.19
5001-10000	0.59	0.42	0.26	0.12
10001-20000	0.69	0.53	0.36	0.21
Above 20,000	0.64	0.44	0.36	0.24
	TTO mean values for disease states			
	State A	**State C**	**State E**	**State F**
0-2000	0.69	0.49	0.15	0.09
2001-5000	0.76	0.63	0.32	-0.02
5001-10000	0.70	0.56	0.14	-0.12
10001-20000	0.78	0.50	0.45	0.16
Above 20,000	0.69	0.23	0.07	-0.12

No significant differences using Kruskal Wallis test of one-way ANOVA F statistic

Table A8.13: Differences in mean values between those with illness and not during valuation (n=117)

	VAS			TTO		
STATES	**ill**	**Not ill**	**Mean difference**	**ill**	**Not ill**	**Mean difference**
State A	0.67	0.66	0.01	0.72	0.75	-0.03
State C	0.48	0.48	0.00	0.46	0.59	-0.13[ab]
State E	0.31	0.32	-0.01	0.13	0.29	-016[a]
State F	0.17	0.18	-0.01	-0.05	0.02	-0.07

a- $p<0.05$ using t test. b – $p<0.05$ using Mann-Whitney test

Table A8.14: Differences between patient and community values for VAS and TTO

	VAS			TTO		
STATES	**patient**	**community**	**Mean difference**	**patient**	**community**	**Mean difference**
State A	0.67	0.66	0.01	0.76	0.72	0.04
State C	0.47	0.50	-0.03	0.54	0.56	-0.02[b]
State E	0.30	0.34	-0.03	0.20	0.27	-0.07
State F	0.18	0.17	0.01	-0.04	0.04	-0.08

b – $p<0.05$ using Mann-Whitney test. No significant difference using t test.

Table A8.15: Differences in mean values for VAS and TTO by ratings of health status in the last two weeks (n=117)

	VAS mean values for disease states			
VAS rating of health status in the last two weeks	**State A** [ab]	**State C** [ab]	**State E**	**State F**
Less than 25	0.62	0.55	0.45	0.19
26-50	0.67	0.49	0.31	0.18
51-75	0.55	0.40	0.28	0.15
75-100	0.70	0.50	0.33	0.19
	TTO mean values for disease states			
	State A	**State C**	**State E**	**State F** [b]
Less than 25	0.91	0.59	0.39	0.54
26-50	0.78	0.56	0.19	-0.21
51-75	0.73	0.51	0.25	0.03
75-100	0.72	0.56	0.25	0.02

a- $p<0.05$ using one-way ANOVA F statistic. b – $p<0.05$ using Kruskal Wallis test

Table A8.16: Differences in mean values for VAS and TTO by ratings of usual health status (n=117)

	VAS mean values for disease states			
VAS rating of usual health status	**State A**	**State C**	**State E** [a]	**State F**
Less than 25	0.55	0.70	0.65	0.27
26-50	0.62	0.58	0.41	0.30
51-75	0.67	0.47	0.32	0.18
75-100	0.67	0.48	0.31	0.17
	TTO mean values for disease states			
	State A	**State C** [a]	**State E**	**State F**
Less than 25	0.96	0.54	0.54	0.46
26-50	0.80	-0.04	0.06	-0.25
51-75	0.81	0.67	0.28	-0.05
75-100	0.71	0.55	0.23	0.01

a- $p<0.05$ using one-way ANOVA F statistic. No significant difference using Kruskal Wallis test

Table A8.17: Socio-economic and demographic characteristics for initial and test re-test valuation samples

	Initial valuation sample (n=117)	Test retest sample (n=60)
Age: (years) Mean (SD) [range]	36.9 (17.0) [15-77]	35.7 (16.4) [15-77]
Gender (%)		
Males	40.2	46.7
Females	59.8	53.3
Marital status (%)		
Single	21.4	28.3
Married	67.5	63.3
Separated / divorced	5.1	5.0
Widow / widower	5.9	3.3
No. of children		
% with children	78.6	70.0
Mean (SD) [range]	4.2 (3.0) [1-15]	4.1 (2.9) [1-10]
Education level (%)		
None	12.8	8.3
Primary	64.9	58.3
Secondary (O and A level)	20.5	30.0
Degree	1.7	3.3
Occupation (%)		
Farmer	72.7	70.0
Teacher	0.9	1.7
Business person	7.7	6.7
Casual labourer	7.7	6.7
Student	9.4	13.3
Other	1.7	1.7
Monthly expenditure ('000 KShs)		
Mean (SD) [range]	8.6 (7.9) [0-37.7]	9.6 (9.1) [0-37.7]

Tests for significant differences used the Mann Whitney test. No significant differences were found

Table A8.18: Proportion reporting illness and VAS ratings of health state (n=60)

	Initial valuation (n=60)	Retest valuation (n=60)	Difference[a]	Spearman's correlation
VAS rating of health state in the last two weeks: (mean (SD) [range]	74.0 (23.6) [10-100]	78.5 (17.9) [30-100]	4.5	0.52**
VAS rating of usual health state: (mean (SD) [range]	84 (17.2) [10-100]	86.9 (13.8) [50-100]	2.9	0.65**
Whether ill in last two weeks (% yes)	41.7	35.0	-6.7	0.37**
Whether ill during valuation (% yes)	26.7	25.0	-1.7	0.37**

[a]Wilcoxon sign rank test did not show any significant differences. ** All correlations significant at p<0.01

Table A8.19: Inter-rater reliability: Practicality (n=60)

	VAS				TTO			
Variables	Initial valuation	Test-retest valuation	Difference	Spearman's correlation	Initial valuation	Test-retest valuation	Difference	Spearman's correlation
Number of times procedure explained (%)								
1	90.0	98.3	8.3[b]	0.40**	96.7	100.0	3.3	n.a
2	5.0	1.7	-3.3		3.3	0.0	-3.3	
3	3.3	0.0	-3.3		0.0	0.0	0.0	
4	1.7	0.0	-1.7		0.0	0.0	0.0	
Ease of using techniques								
Very easy	5.0	1.7	-3.3	0.37**	1.7	0.0	-1.7	0.18
Easy	55.0	56.7	1.7		45.0	33.3	-11.7	
Fairly easy	26.7	35.0	8.3		25.0	45.0	20.0	
Difficult	11.7	6.7	-5.0		28.3	21.7	-6.6	
Very difficult	1.7	0.0	-1.7		0.0	0.0	0.0	
Ease of VAS compared to TTO and TTO compared to VAS								
Easy	58.3	65.0	6.7	0.18	36.7	25.0	-11.7	0.37**
Somewhat easy	36.7	30.0	-6.7		43.3	53.3	10.0	
Difficult	5.0	5.0	0.0		20.0	21.7	1.7	
Extent of thinking required to use technique								
Very hard	10.0	5.0	-5.0	0.58***	8.3	11.7	3.4	0.211
Hard	25.0	18.3	-6.7		50.0	51.7	1.7	
A little	55.0	68.3	13.3		36.7	35.0	-1.7	
Very little	10.0	8.3	-1.7		5.0	1.7	-3.3	
							0	
Preferred technique	60.0	71.7	11.7		40.0	28.3	-11.7	0.17[a]

** p<0.01 *** p<0.001 n.a – all explained for 1 time at retest. Therefore no correlation coefficient computed

[a] correlation for preference between VAS and TTO [b] shows that the difference between test and retest is significant at p<0 05 using the Wilcoxon sign rank.

Table A8.20: Mean VAS and TTO values between initial and retest valuation (n=60)

		VAS		TTO	
STATES		Initial	Retest	initial	retest
State A	Mean	0.64	0.67	0.80	0.78
	Standard deviation	0.16	0.15	0.18	0.20
	Range	0.4-0.9	0.4-0.9	0.04-0.96	0.04-1
	95% C.I.	0.60-0.68	0.63-0.71	0.75-0.85	0.72-0.83
State C	Mean	0.46	0.49	0.63	0.61
	Standard deviation	0.14	0.13	0.35	0.32
	Range	0.14-0.8	0.18-0.8	-0.96-1	-0.96-0.96
	95% C.I.	0.43-0.50	0.45-0.52	0.54-0.72	0.53-0.69
State E	Mean	0.32	0.33	0.30	0.35
	Standard deviation	0.14	0.11	0.50	0.29
	Range	0.1-0.7	0.1-0.7	-0.96-0.96	-0.79-0.96
	95% C.I.	0.29-0.36	0.30-0.36	0.17-0.43	0.28-0.43
State F	Mean	0.19	0.21	-0.02	0.23
	Standard deviation	0.13	0.10	0.57	0.41
	Range	0-0.6	0.08-0.6	-0.96-0.96	-0.96-0.86
	95% C.I.	0.15-0.22	0.19-0.24	-0.17-0.13	0.13-0.34

APPENDIX 9.1: ISSUES IN ASSESSMENT OF HRQL IN CHILDREN

To date, measurement and valuation of HRQL has tended to focus on adult populations and rarely on children (Bullinger, 1997: Pal, 1996). Where HRQL of children has been measured and valued, proxies (parents, caregivers, and professionals) have been used (Rosenbaum and Saigal, 1996: Pal, 1996). Issues surrounding measurement and valuation of HRQL of children by children relate to level of cognitive capacity, their perception and understanding of recall period, age, the problem of whose perspectives to use in assessing HRQL, effects of development on children's function (Feeny, 1999: Rosenbaum and Saigal, 1996: Wright, 1996). Other issues involve which aspects or domain of function to include as life experiences and daily activities of children and adults differ substantially (Rosenbaum and Saigal, 1996). For example, Manificat and Dazord (1998) found that whilst some domains are similar to those found in adult questionnaires, others are quite specific to childhood such as relationships, external events and feelings. An integral component of children's health has been defined in terms of the ability to fulfil age-related activities including physical, emotional and social activities. In addition, developmental stage, diseases perception and coping styles have to be considered (Bullinger, 1997). In terms of Herdman et al's (1998) model, these are issues related to conceptual, operational and functional equivalence.

It has been argued that parents do not accurately perceive children's HRQL and that information must be obtained from children themselves (Juniper et al. 1997: Apajasalo et al. 1996: Eiser et al. 1995 and Eiser, 1997). Juniper et al (1997) show that children (7-17 years) provide very reliable data but note that they needed over grade six reading skills to complete the SG and over grade two for the feeling thermometer. Children over the age of 7 years can complete the HUI 2 and 3 questionnaire consistently. Feeny (1999) argues that children (as young as 7 years) are quite capable of responding on their own behalf and they can provide reliable, valid and useful information. Feeny (1999) and Landgraf and Abetz (1996) report that responses by younger children match those of blinded clinicians than their parent's rating.

Other studies supporting "asking the children themselves" include Bullinger (1997) who notes that although expert rating rather than parent's or clinical ratings have been favored, the essence of quality of life is self rating method, thereby supporting the view that children should rate their own health whenever possible. Landgraf and Abetz (1996) suggest that although it has been suggested that children as young as 5 years can provide empirically reliable reports on pain and over the counter medication use, a more conservative estimate would be 9 or 10 years of age for more subjective subjects such as behavior or self esteem.

These studies suggest that children can rate and provide values for their own health states, although certain issues like age, cognitive capabilities, developmental effects should be taken into consideration while observing maintenance of various types of equivalence. Due to the special considerations required in measuring and valuing children's health, the issue was not considered in this thesis.