CENTRE FOR HEALTH
PROGRAM EVALUATION

The Australian Multi-attribute Utility (AMAU) construction and Initial Evaluation

Dr Graeme Hawthorne
Senior Research Fellow, Program Evaluation Unit, CHPE

Dr Richard Osbourne
Research Fellow, Health Economics Unit, CHPE

Ms Helen McNeil
Research Fellow, Health Economics Unit, CHPE

Professor Jeff Richardson
Director, Health Economics Unit, CHPE

CENTRE PROFILE

The Centre for Health Program Evaluation (CHPE) is a research and teaching organisation established in 1990 to:

- undertake academic and applied research into health programs, health systems and current policy issues;

- develop appropriate evaluation methodologies; and

- promote the teaching of health economics and health program evaluation, in order to increase the supply of trained specialists and to improve the level of understanding in the health community.

The Centre comprises two independent research units, the Health Economics Unit (HEU) which is part of the Faculty of Business and Economics at Monash University, and the Program Evaluation Unit (PEU) which is part of the Department of General Practice and Public Health at The University of Melbourne. The two units undertake their own individual work programs as well as collaborative research and teaching activities.

PUBLICATIONS

The views expressed in Centre publications are those of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of publications are encouraged to contact the author(s) with comments, criticisms and suggestions.

A list of the Centre's papers is provided inside the back cover. Further information and copies of the papers may be obtained by contacting:

The Co-ordinator
Centre for Health Program Evaluation
PO Box 477
West Heidelberg Vic 3081, Australia
Telephone + 61 3 9496 4433/4434   **Facsimile** + 61 3 9496 4424
E-mail CHPE@BusEco.monash.edu.au

ABSTRACT

This paper presents further work on the development of the Australian Multi-attribute Utility (AMAU) instrument, the first steps of which were described in Hawthorne and Richardson (1996). The purpose of the instrument is to assist with the measurement of health-related quality of life (HRQoL); that is, it is designed for use in cost-utility analysis.

The present paper reports the survey and statistical methods adopted for the construction of the descriptive system. It presents a descriptive system which consists of five dimensions and 15 items. This structure is consistent with the hypothesised structure and dimensions of HRQoL.

The psychometric properties of the instrument are reported, including preliminary indicators of its capabilities. The results suggest the AMAU has the ability to distinguish levels of health evident in hospital and community cases, between age groups and educational achievement. In addition, interviews with subjects show it is very easy to understand and complete with minimal difficulty. The AMAU is thus a short and convenient measure of HRQoL.

In summary, the descriptive system outlined in this paper appears suitable for assessing HRQoL in Australian populations; and as a MAU instrument for calculating QALYs for economic and health program evaluation.

TABLE OF CONTENTS

4   Discussion

Summary and future work

References

Appendix 1: The AMAU instrument

THE AUSTRALIAN MULTI-ATTRIBUTE UTILITY INSTRUMENT (AMAU):

Construction and Initial Evaluation

1 Introduction:

1.1 Background

Before the development of cost utility analysis, economic evaluation of health services either ignored quality of life (QoL) or treated it as an 'intangible' that could be noted and described but not quantified or included as an integral part of the health outcome. Cost utility analysis (CUA) has attempted to overcome this deficit by adopting the quality adjusted life year (QALY) as the unit of output for health benefits in cost effectiveness studies.

While there have been numerous criticisms of the concept and its measurement (Burrows and Brown 1993; Richardson 1991) its rationale is simple. Life years are weighted by an index of utility—a preference weight—which acts as an exchange rate between the quantity and quality of life. This implies that a "strong" interval property must be met by the utility index. Not only must a given reduction in its value mean the same across the entire range of the index but it must mean the same as an equivalent reduction in the quantity of life. Thus, for example, a 0.2 reduction in the utility index from 0.8 to 0.6 must be equivalent to a reduction from 0.5 to 0.3, and this must be equivalent to a corresponding reduction in life expectancy.

Achieving this property is highly problematic, but in economic evaluation a comparison of changes in QoL and life expectancy cannot be avoided. The only way to achieve this property is through using of one of the preference based scaling techniques; *viz*, time trade-off (TTO), standard gamble (SG) or the person trade-off (PTO). Each asks respondents to compare

QoL and life expectancy and to trade-off one against the other. The validity of the scale values obtained from the use of these techniques cannot be determined by direct comparison with actual behaviour: people cannot be easily observed trading QoL against life expectancy, and so the validity of this depends primarily upon the belief that people may make meaningful judgments about the trade-off. Conventional forms of validity must still apply, however, and these include content (does the content of the instrument adequately reflect that of the theoretical universe upon which it is premised?), criterion (how well does the measurement match that from other instruments purporting to measure the same thing?) and construct validity (does the instrument provide scores from which inferences about the universe can be made?). Few validity tests of MAU instruments have been reported in the literature.

The utility values of a health state may be obtained directly by the application of a scaling technique to a description of the health state, or indirectly by the use of a multi-attribute utility (MAU) instrument. Such instruments are based upon a 'descriptive system' which is a set of 'items' that describe different levels of some aspect of health (usually referred to as a health 'dimension', such as mobility, pain or depression). The full set of these items define and limit the health states that can be described by the instrument. Corresponding with the descriptive system is a set of utility values or a valuation formula to determine the utility value of each of the health states defined by the descriptive system. Several MAU instruments are readily available; *viz*, the Rosser-Kind instrument, the Quality of Well-Being (QWB), the Health Utility Index Mark I, II & III (HUI), the 15D and the EUROQoL (for details see Table 12).

While it is beyond the scope of the present paper to offer a full critique of existing instruments it is the contention here—and the motivation of the present project—that these instruments have shortcomings. Three have particularly simple, and therefore rather insensitive, descriptive systems; the EUROQoL, the original Rosser-Kind and the original McMaster instruments (Drummond, Stoddart et al. 1987; Williams 1995). The validity of at least two of the instruments is open to question; the QWB and HUI Mark I (Nord, Richardson et al. 1993). While the HUI Mark II is a generic instrument it was designed for a particular task and population, and the HUI Mark III is intentionally focussed primarily upon bodily disability. While purporting to produce utility scores for the calculation of QALY values, several of the instruments were constructed without the use of preference scaling techniques; the 15D, Rosser/Kind, and QWB. (Drummond, Stoddart et al. 1987; Rosser 1993; Sintonen 1994; Feeny, Furlong et al. 1995; MVHGroup 1995; Sintonen 1995; Williams 1995; Feeny, Torrance et al. 1996.)

In addition, the validity of the modelling procedures used to obtain utility scores for each of the multi-attribute health states is largely untested (Sintonen 1995). These issues indicate the present situation regarding the measurement of QALYs for economic evaluation is unsatisfactory.

An acceptable descriptive system must be capable of describing and discriminating between health states before and after the interventions that are to be evaluated with the instrument. Given the limitations above, the objective of the present project is to construct an instrument—including a sound descriptive system—that is capable of evaluating a very wide range of interventions.

1.2   The Australian MAU project

The present project involves the construction of an Australian MAU-QALY instrument for use in evaluation and cost-utility analyses aimed at assisting with decisions concerning the allocation of resources between competing health related interventions. More specifically, it is developing an instrument that can be used in the evaluation of a wide range of programs for the treatment of acute and chronic conditions, ambulatory and institutional care, public health, health prevention and promotion programs. The project is not simply to replicate existing instruments, but to construct a descriptive system that has strong psychometric properties and then to scale it using values from a random cross-section of the Australian population. The aim is to construct an instrument without the shortcomings referred to above.

The work reported in this paper continues that described previously (Hawthorne and Richardson 1996); *viz*, the development of the multi-attribute utility instrument. The instrument is based upon solid theoretical principles, it will be standardised and validated for the general Australian population, and it will be applicable to specific groups within the community.

The validity of the HRQoL measurement, the utility scores and the resulting QALY estimates all rest on the quality of the *descriptive system* upon which these estimates are based. As such, descriptive systems must meet the standard psychometric criteria for validity (as outlined above) and reliability.

The reason for this is that health-related 'quality of life' is a psychometric construct related to 'health'. Essentially, 'health' itself is a psychometric construct in that a person's perception of the value of any given health-state is a subjective estimate based on their core constructs, attitudes, beliefs and values. Defining 'health', though, is problematic as it may be based on

concepts as inclusive as the Old English *hælth*, as narrow as the prevailing bio-medical perspective or as political as the social construct paradigm. As the objective here was to construct an instrument relevant to a broad range of interventions—including health promotion—a very broad concept of health was adopted; a concept based upon the World Health Organisation's (WHO) notion of handicap (see Hawthorne and Richardson 1996). Specifically, the WHO definition of health was adopted: that health is a state of optimum physical, mental and social well-being and not merely the absence of disease or infirmity (WHO 1958). This implies that health embraces all aspects of life as exemplified by Ware's phrase that health connotes 'completeness' (Ware 1987). In terms of defining an individual's health state it follows that the so-called 'gold standard' for any individual is when they are in this optimum state, and that any deviation from this constitutes a less-than-optimum health state. Since less-than-optimum health impacts upon individuals' everyday lives, in this study ideal HRQoL was defined as *living without handicap attributable to health status*. Consequent upon this definition, the measurement of HRQoL focused upon departures from this idealised health state, specifically where ill-health restricted a person's activities through 'handicap'.

Development of psychometric instruments—including descriptive systems to measure HRQoL as defined—is a difficult and costly exercise. Best practise in their construction involves: first, development of a clear theoretical framework and an item pool (a list of potential questions) based on this framework; second, ensuring the item pool covers the breadth of potential health states of interest (ie a wide variety of health states); third, testing these items in populations that are representative of the populations the instrument will be applied to; and finally, through a combination of logic and iterative factor analysis techniques, concluding with the development of a parsimonious instrument in which each item contributes uniquely to the measurement of HRQoL.

In addition, and complementary to, these psychometric requirements, Ware argued HRQoL instruments should be based on a clear concept of health, they should embrace the best and worst health states—ie they should avoid being focussed on the disease end of the health spectrum—and they should be able to differentiate between hospital patients and community cases (Ware 1987).

The above schema has been followed in the construction of the AMAU descriptive system; the methodology and results of this work are described in this paper. The next steps, which are currently in progress, involve weighting the descriptive system and obtaining estimates of its validity and reliability.

In summary, HRQoL was defined through a review of the literature and consultation with health professionals. It was hypothesised that HRQoL comprised 5 major dimensions (Table 1). A sixty-one item pool was developed and administered to 143 people in hospital and 112 people from the general community. Through logical and statistical evaluation of items and an iterative factor analytic approach, a 5-dimension 15-item AMAU (Appendix 1) was constructed consistent with the original hypothesis.

Table 1 Proposed MAU dimensions and elements

Dimension	Elements
Illness	Pain
	Receiving medical treatment
	Reliance of medical aids
Independent living	Activities of daily living
	Bodily care
	Safety and security
Physical ability	Mobility
	Physical performance
	Rest and fatigue
	Sensory function
	Sexual relationship
Psychological state	Anxiety/depression
	Cognitive ability
	Emotional fulfilment
	Memory
Social interaction	Communication
	Family role
	Intimacy
	Social function
	Work satisfaction

2   Methods:

2.1   Summary: process of development

The program to develop an AMAU was initiated in 1993 following recognition of the inadequacy of existing multi-attribute utility instruments (Hawthorne and Richardson 1996). The development of a theoretical descriptive system began with a literature review, including a detailed analysis of existing QoL and MAU instruments. This led to the development of a HRQoL model comprising 5 dimensions. An item pool was developed based on this theoretical model. Potential items were generated from existing instruments, through focus groups and face-to-face interviews with health professionals. The resulting initial item pool was reassessed by health professionals to ensure items reflected the range of health states encountered in hospital settings, ie from best to worst health states actually encountered in clinical practice. A final pool of 61 items representing the 5 dimensions were selected for trial in two groups of subjects; those during an admission to a hospital and subjects randomly recruited from the general community.

The list below (a) to (i) summarises this developmental process. Sections (a) to (e) were more fully described in the preliminary report (Hawthorne and Richardson 1996) and are not discussed further in this report.

1   Literature review

2   Construction of a model of HRQoL

3   Generation of draft items

4   Rewriting and clarification of items through interviews and focus groups with medical personnel

5   Sixty-one items were selected for trial

6   Drawing of subjects: random selection of hospital patients and community cases

7   Interview

8   Item reduction

9   Construction of a parsimonious instrument with a logical factor structure and sound psychometric properties.

In the survey—steps (f) and (g)— subjects were asked to describe their own health states using the 61 items. This provided the information required for an analysis of item sensitivity

and the factor structure of the instrument (ie the groupings of items which form a coherent dimension).

2.2 Subject recruitment and interview procedure

Successful instrument construction relies upon data being gathered from a sample drawn from the population with which the final instrument will be used. Because of the populations that the AMAU will be applied to in future research the instrument must be based upon values elicited from hospital and community samples. To ensure the construction sample reflected these populations and contained the widest possible range of health states the sample was drawn from a hospital and the general community, stratified by age and gender (Table 2). Additionally, assessing these two groups enabled identification of draft items that discriminated between 'ill' people (hospital patients) and 'healthy' people (community cases).

To ensure representativeness, adolescents were also recruited as a comparison group.

2.2.1   Hospital patients

Hospital patients were identified via a listwise selection procedure at St Vincent's Hospital, a major public teaching hospital in Melbourne. This involved the interviewer working through hospital ward lists (in consultation with the charge nurse) and approaching eligible patients. Patients were eligible if they fulfilled the following criteria:

- able to understand sufficient English, ie they were able to read and comprehend the questionnaire;

- willing to sign a consent form;

- aged 18 years or older (no upper limit); and if they were

- not cognitively impaired due to their health condition or treatment (judged by charge nurse on duty).

Eligible patients were approached after admittance to the ward or during their recovery in the ward. The following style of introductory statement was presented to each patient:

> *A health survey is being conducted by Monash University with the cooperation of the hospital. Participation is voluntary. Would you like to hear more about what may be involved should you feel like participating?*

Interested patients were given an explanation of what was involved in the survey; this included a five minute interview about their background and a questionnaire that would take about half an hour to complete. Patients who expressed an interest in taking part were requested to complete a consent form.

The first part of data collection involved a semi-formal, but structured, interview regarding their demographic background. The second part involved completing the 61-item questionnaire. Once the interviewer was confident the subject understood the questions and the answering procedure, the subject was left to complete the questionnaire at their own pace. Questionnaires were collected later that day and the questionnaire was checked for completeness. Any difficulties the subject may have encountered were discussed and noted. When collection was impossible on the administration day (eg due to treatment) questionnaires were collected the following day.

2.2.2   Community cases

Community cases were selected through random selection from the Melbourne telephone directory. Potential subjects were sent a letter informing them of the study and that in a few days they would receive a telephone call to discuss their willingness to participate.

Eligibility criteria were the same as for hospital patients (above). The potential inconvenience for community cases was minimised through conducting interviews at their home at a time that was convenient to them. In a similar manner to the patients, the questionnaire was administered to each subject by an interviewer. Once the interviewer was confident the subject was able to fill in the questionnaire correctly, they were left to complete it in their own time. The interviewer collected the questionnaire later that day, or the subject posted it back the Centre for Health Program Evaluation (CHPE). If the interviewer had doubts as to whether the subject would reliably complete the whole questionnaire or might not post it back, the interviewer remained with the subject until the questionnaire was completed and retrieved it at the end of the interview.

2.2.3   Adolescents

A sample of adolescents also completed the 61 item pool as a means of estimating preliminary performance indicators after the initial instrument was constructed (see Section 3.5).

Adolescents were selected and interviewed with the cooperation of Ms Caroline Coffey, Centre for Adolescent Health, Royal Children's Hospital. One hundred and fifty Year 10 students from six Melbourne secondary colleges completed the questionnaire. Adolescents completed the questionnaire during a teaching period and due to this time constraint, the item *difficulty* question was omitted (see below).

2.2.4   Questionnaire format

The 61 trial items, each with Guttman or Likert response scales, were presented in random order (see Figure 1 for a Guttman example). After each question, subjects were asked to score how easy or difficult they found the question (scores ranged from 1 = very easy, to 7 = very difficult). Following completion of the QoL items subjects were asked to rate the importance[1] of each domain (scores ranged from 1 = not important at all, to 7 = extremely important).

Subjects were encouraged to clarify any difficulties during the interview (community cases and adolescents) or at the end when the interviewer returned to collect the questionnaire (hospital patients). Very few subjects reported difficulties in understanding the questionnaire (see below). Some subjects complained of the repetitious nature of some items; a problem that is normally encountered during *item reduction* phases of instrument development.

---

[1]Importance ratings were administered to estimate the importance of each dimension of HRQoL. The data for this procedure is not presented in this report and will be presented in a later report.

Figure 1   Example Item and difficulty rating.

6	**Do I need any help looking after myself?**
A.	I need no help at all.
B.	Occasionally I need some help with looking after myself.
C.	I need help with the more difficult personal care tasks.
D.	I need daily help with most or all personal care tasks.

How easy or difficult was it to answer the above question?

very easy		easy		difficult		very difficult
1	2	3	4	5	6	7
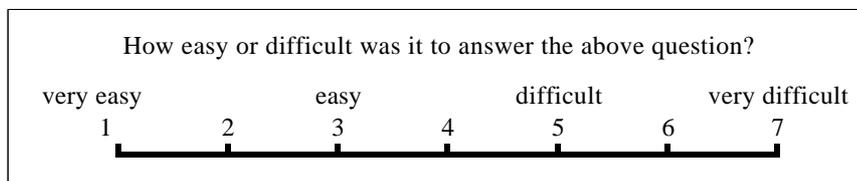
2.3   Statistical analysis

2.3.1   Sample size

There is a debate in the QoL literature over whose values ought to be used when weighting items within instruments: some have argued weights ought to reflect the experiences of patients whilst others have argued those of the general population ought to be used.

In constructing the AMAU, it was determined that this problem would be addressed through collecting data from both patients and the general population in order to establish whether there is a clinically significant difference (based on Drummond and Davies 1991). The null

$$H_0 : \overline{X}_{patients} = \overline{X}_{population}$$

hypothesis was thus:

To estimate appropriate sample size to test this hypothesis, four parameters were established:

- The 'size' of the test (α); here set at the conventional p=0.05.
- The 'power' of the test ($1-\beta$); again set at the conventional P=0.20.

- The 'effect' of the variable of interest; the clinically significant difference was set at $\Delta= 5$pts
- The 'variability' ; in this case it was estimated to be $S_x =0.1470$
- The unit of analysis; this was the individual subject.

The computations for Δ and S_x relied upon conventional sample size estimates: these suggested an estimated 107 subjects would be needed in each group: a total of 214 subjects.

The estimate of the S_x was based on the NCHPE lithotripsy study (Cook and Richardson 1993). This study reported 'utility health state values' for patients with gallstones. Although no population estimates were available, the upper patient estimate was used as these subjects were not subject to severe discomfort. Since the scale was constrained to 0–1—as is intended with the AMAU—it may be that there was a ceiling effect at the upper scale limit, ie scores above 0.90 may be artificially constrained where the scale is applied to a general population. The highest value, therefore, from this study was accepted as the general population mean estimate for calculation of sample size.

- The mean patient utility value scores for the upper 'utility health state values' ($\overline{X}_U = 0.90,\ 95\ \%\ C\ I : 0.87 \_ 0.93$), the closest to upper ($\overline{X}_C = 0.88,\ 95\ \%\ C\ I : 0.86 \_ 0.90$) and the lowest health state values ($\overline{X}_L = 0.68,\ 95\ \%\ C\ I : 0.63 \_ 0.74$)
- Number of cases: n = 96

From these data the standard errors and standard deviations were estimated. The critical unknown factor related to the estimated smallest *clinical* difference which should be detected: clearly a difference as great as that between the lithotripsy upper and lower values was too great to be used, and likewise accepting the upper and closest estimates was equally undesirable. Instead, a 20% difference in scores for TTO to assume clinical significance was adopted (Drummond and Davies 1991). This equated with a 5-point difference in scores.

When the decision was made that this difference in scores should be detected, assuming a population mean of 0.90 and that patients' mean score would be lower, and that the standard deviations were constant, it was computed that 107 subjects would be needed in each of the hospital and community samples.

2.3.2 Examination of the data

Tests of normal distribution of data were estimated via frequency plots, kurtosis, skewness and through the Kolmogorov-Smirnov test.

Differences between groups were examined via either Student's t-test (for normally distributed data) or the Mann-Whitney U-test (non-normally distributed data). Correlations between variables were investigated via Pearson's correlation coefficient (interval data) or Spearman's correlation coefficient (ranked data).

Estimation of whether items were measuring a single latent construct (HRQoL) was done via a principal components analysis. Existence of underlying factors leading to the construction of the multidimensional HRQoL scale was via an iterative factor analysis involving both non-rotated and varimax rotations. The reliability of the instrument was assessed via Cronbach's α.

Analyses were carried out using SPSS for Windows (Norusis 1993).

3 Results

Data were collected by two experienced interviewers from July to December 1995. Data were entered into and cleaned in Microsoft Excel and analysed using SPSS for Windows (Norusis 1993).

3.1 Sample description

3.1.1 Sample size and recruitment

The total number of subjects recruited for this study was 255 adults (143 hospital patients and 112 community cases). A comparison group of 150 adolescents was also recruited. The minimum sample required for the two adult populations was 107 (see above).

In the hospital sample a total of 179 eligible patients were approached, 161 patients consented and participated and 19 patients refused to participate. This indicated a recruitment rate of 90%. In 18 cases (11%) the data were lost to due to poor completion.

In the community sample—randomly selected from the Melbourne telephone directory—a total of 287 people were initially drawn (ie the sample size plus replacement; the sample was overdrawn in the expectation that potential subjects would be non-contactable). Of these, 79 were uncontactable, 30 were ineligible, 138 consented to participate and 40 refused. This indicated a recruitment rate of 78%. In 26 cases, the data were lost to follow-up (19%).

3.1.2 Gender and age

The hospital and community samples were stratified by gender and age. This stratification was used because HRQoL is known to vary between age groups and gender. Table 2 shows the gender and age distribution indicating an even distribution with no cells having a disproportionate number of subjects.

Table 2 Age of subjects by group and sex

	Hospital				Community		
	Female	Male	Total		Female	Male	Total
29 and younger	11	15	26		12	8	20
30 to 39	13	10	23		9	11	20
40 to 49	11	8	19		15	8	23
50 to 59	10	13	23		7	11	18
60 to 69	14	10	24		9	8	17
70 and older	14	14	28		9	5	14
Total	73	70	143		61	51	112

$\chi^2 = 4.1$, DF = 5, p = ns

3.1.3   Socioeconomic status indicators (educational level and employment status)

Education

As shown in Table 3 below, hospital patients had a lower educational level when compared with community respondents. Fifty-three percent of patients had primary education only compared with 27% of community cases. Only 12% of patients had a university education compared with 33% of the community cases.

Table 3 Educational level

	Hospital			Community	
	Count	%		Count	%
No education	2	1%		0	0%
Primary only	73	53%		30	27%
Secondary (completed year 12)	29	21%		20	18%
Trade or TAFE	19	14%		25	22%
University / postgraduate	16	12%		37	33%

	Count	%		112	100%
Total	139	100%		112	100%

<div align="right">

χ^2 = 28.1, DF =4, *p* <0.001

Note: 4 missing cases

</div>

Employment

Hospital patients were less likely to be employed (31%) when compared with community respondents (57%). In addition, more patients were on sickness/disability pensions (14%) than were community cases (2%), as might be expected. These data are displayed in Table 4.

Table 4 Employment status

	Hospital		Community	
	Count	%	Count	%
working	44	31%	64	57%
unemployed	9	6%	3	3%
retired	45	32%	22	20%
keeping house	16	11%	18	16%
invalid/sickness pension	20	14%	2	2%
student	6	4%	3	3%
other	3	2%	0	0%
Total	143	100%	112	100%

<div align="right">

χ^2 = 30.1, DF = 6, *p* <0.001

</div>

3.1.4 Country of birth

Table 5 reports the distribution of the country of birth of hospital and community cases. Hospital patients were significantly less likely to have been born in Australia (22%) compared with community respondents (33%).

Table 5 Country of birth

	Hospital	Community	Total
Non-Australian	31	37	68
Australian	112	75	187
Total	143	112	255

$\chi^2 = 4.14$, DF = 1, $p = 0.042$

3.1.5   Diagnosis Related Group distributions

Information was obtained from the hospital medical records on the patient's diagnosis related group (DRG). In Table 6 it can be observed that the majority of patients (68%) were from 5 DRG's: circulatory (17.5%), digestive (14%), urinary/kidney (14%), musculoskeletal (12.6%) or nervous system (10.5%) disorders.

Table 6 DRG distributions (hospital patients)

DRG group	Count	%
Circulatory system	25	18%
Digestive system	20	14%
Kidney and urinary systems	20	14%
Musculoskeletal & connective tissue	18	13%
Nervous system	15	11%
Hepatobilary system and pancreas	10	7%
Respiratory system	7	5%
Skin, subcutaneous and breast	7	5%
Endocrine, nutritional, metabolic etc	4	3%
Myleoproliferative, poorly differentiated neoplasms	4	3%
Injuries, poisonings, toxic effects of drugs	4	3%
Mental disorders	3	2%
Factors influencing health states, other contact with health	3	2%
Male reproductive system	1	1%
Female reproductive system	1	1%
Infections, parasitic	1	1%
Total	143	100%

3.2   Easy/Difficult rating of items and missing values

To obtain an estimate of the ease or difficulty of individual items, subjects were required to rate each item immediately after answering. The rating system using a Likert scale ranged from 'very easy' (1) to 'very difficult' (7) (see Figure 1). When a subject rated a question at 5 ('difficult') or above the interviewer probed the subject to determine why they found the item difficult to answer. Notes were taken and collated for each item. The average difficulty rating across all items was 1.9 ± 0.23 (mean ± standard deviation) indicating that most subjects found the items easy to answer. One item stood out from the others (item MAU5.3: score = 2.5) which referred to difficulty of breathing during physical activity. Comments from subjects indicated this was the only item inapplicable to their situation. This item also had the highest number of missing values (8.4%).

Generally, the stated reasons for items being difficult fell into two categories:

(a)   the wording of the item was difficult to understand; the question was long and involved, or the question did not apply to their life situation; or

(b)   the question was difficult because the subject reported experiencing a *low quality of life* and in order to answer the question they were forced to acknowledge or think about this.

These difficulties were rare as indicated by the response rate. The mean valid response rate for all items was 98.7%. As indicated above only one item stood out as a "problem item" (item MAU5.3) and this was deleted.

In general, subjects found the items easy to respond to. The difficulty rating, in addition to providing information about subjects' subjective responses, provided valuable information for the ensuing *item selection* stage of instrument construction.

3.3   Instrument construction

3.3.1   Item selection

Item selection procedures are described in the previous paper (Hawthorne and Richardson 1996) and are only summarised here. The criteria are listed below. An additional criterion was whether an item was able to discriminate between hospital and community cases. When an item did not fulfil a criterion, it was deleted. However, this process was tempered by a logical analysis to check whether a particular item possessed characteristics that overrode removal through statistical criteria (see below).

3.3.2 Stages in the item selection procedure

Stage 1

- Item sensitivity: items with a range less than 100% were discarded. The only exception to this was Item 6.4 relating to hearing. It was retained on the grounds that the sample was drawn through telephone contact which would have precluded anyone whose hearing was greatly impaired.

- Item spread: items with a standard deviation less than 0.50 were discarded.

- loading on the principal factor during principal components analysis: items loading less than 0.20 were discarded.

Stage 2

- following re-analysis of the remaining item pool, items loading on the principal factor during principal components analysis less than 0.30 were discarded.

- item-rest-of-test (IRT) score: items scoring less than 0.20 were discarded.

Stage 3

- for each element of each dimension (see Table 1) 4 items had been initially developed. In this stage repetitious items with the poorest features were removed based on item sensitivity, missing values, and highest degree of difficulty. An additional criterion was the ability of an item to significantly discriminate between hospital and community cases.

Stage 4

- the remaining items were pooled and subjected to an iterative two stage factor analysis (principal components and varimax) and to further reliability analysis. Items were excluded it they failed to load (≥ 0.3) on the principal factor during principal component analysis, loaded (≥ 0.3) on more than one factor following a varimax rotation, or obtained a low item rest-of-test r(it) score (< 0.20) or high r(it) score (>0.70). These procedures were adopted to ensure that items covering broad content areas and with a wide range of responses were retained. This ensured that each item made a unique contribution to the instrument and loaded on a logical factor dimension (scale).

- Items which loaded on a factor and were not plausibly related to the logical scale construct were also discarded.

Stage 5

- stages 3 and 4 were repeated iteratively until the most parsimonious solution was derived, which explained the greatest cumulative variance.

- this led to the development of a model consistent with the hypothesised structure of HRQoL (five dimensions; eigenvalue >1.0) where each factor dimension (scale) was orthogonal to the other factor dimensions (scales) as determined by varimax rotation.

- each unique factor was then subjected to principal components analysis with the objective of determining its internal structure. Where more than one factor was identified, items were reviewed and removed prior to the revised scale being returned for further pooled analysis as described in stage 4.

- the internal reliability of the instrument, and of each scale separately, was assessed at each step through Cronbach's α.

3.3.3   Instrument characteristics

The five steps above resulted in a model with 5 factors (graphically illustrated by the scree plot in Figure 2) which explained a cumulative variance of 64.5%. The estimated reliability coefficient (internal consistency) was excellent ($\alpha = 0.80$). A summary table of the final analysis is presented in Table 7.
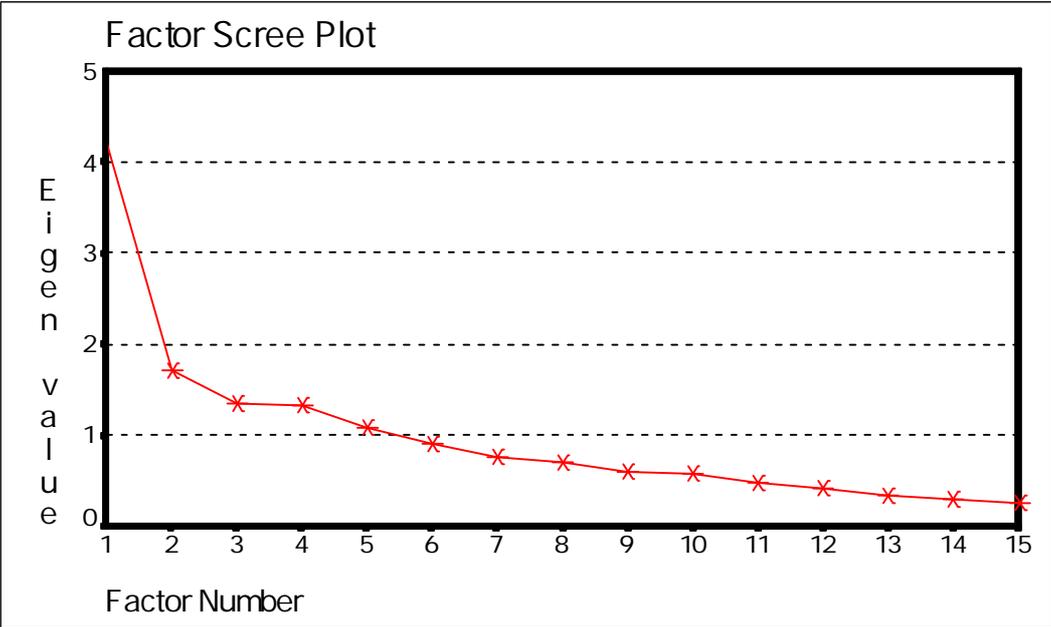
Figure 2 Factor structure (scree) plot of the AMAU.

Table 7. AMAU instrument: factor structure of the descriptive system

Scale and item content	Principal components _analysis			Factor analysis_ (varimax rotation)					Reliability _analysis	
	PC	Eigen value	% _variance	1	2	3	4	5	r(it)	alpha
Scale 1 Illness										
1 Drug use	.65			.85					.54	
2 Medical aids use	.70			.85					.60	
3 Medical treatment	.70	4.2	28.0%	.84					.59	
Scale 2 Independent living										
4 Self care	.61				.87				.46	
5 Activities of daily living	.67				.69				.53	
6 Mobility	.62	1.7	11.5%		.77				.48	
Scale 3 Social Relationships										
7 Social relationships	.33					.83			.26	
8 Sex	.47					.61			.38	
9 Family role	.63	1.3	8.9%		.42	.65			.52	
Scale 4 Physical senses										
10 Sight	.38						.68		.30	
11 Hearing	.30						.78		.23	
12 Communication	.31	1.3	8.9%				.66	.37	.26	
Scale 5 Psychological well-being										
13 Sleep	.40							.72	.32	
14 Anxiety	.48					.43		.62	.42	
15 Pain	.34	1.1	7.3%					.58	.28	
Whole instrument summary			64.6%							.80

Each of the five factors comprised three items. In all cases, items loaded ≥ 0.30 on one factor and only three items had significant loadings on one other factor. The separate analyses of each factor—hereafter referred to as a 'scale'—are given in Table 8. Principal component analysis of each scale revealed the presence of just one factor per scale—suggesting unidimensionality—consistent with the combined pooled varimax analysis (Table 7). The reliability of the five scales ranged from $\alpha = 0.87$ (Scale 1: Illness) to $\alpha = 0.52$ (Scale 5: Psychological well-being).

The factorial structures and reliability estimates indicated that each scale had excellent psychometric properties.

Table 8 AMAU instrument: structure of scales.

Scale and item content	Principal components _analysis			Reliability _analysis	
	PC	Eigen value	% _variance	r(it)	alpha
Scale 1 Illness					
1 Drug use	.89			.74	
2 Medical aids use	.91			.79	
3 Medical treatment	.88	2.4	79.4%	.73	.87
Scale 2 Independent living					
4 Self care	.86			.64	
5 Activities of daily living	.82			.58	
6 Mobility	.78	2.0	66.9%	.52	.75
Scale 3 Social Relationships					
7 Social relationships	.73			.33	
8 Sex	.67			.32	
9 Family role	.80	1.6	53.7%	.42	.57
Scale 4 Physical senses					
10 Sight	.73			.36	
11 Hearing	.76			.39	
12 Communication	.68	1.6	52.4%	.32	.54
Scale 5 Psychological well-being					
13 Sleep	.79			.41	
14 Anxiety	.69			.33	
15 Pain	.66	1.5	51.4%	.30	.52
Whole instrument summary			64.6%		.80

While factor analysis identifies latent factors within groups of variables, it does not provide any interpretation of the meaning of each factor. An accepted practise in psychometrics is to

ascertain the primary theme of a factor through examination of the content of the pivotal items, ie those items with the highest loadings. Where the pivotal items' content is consistent with the hypothesised conceptual structure, it provides evidence that the scale—comprising those items with their highest loading on the factor—validly represents the theoretical model describing the concept under investigation.

The pivotal items resulting from principal components analysis were examined in the above analysis (Tables 7 & 8). The brief descriptions appearing at the heading of each factor (Scales 1 to 5) were based on the interpretation of the pivotal items.

3.4 Congruence of factor analytic model with hypothesised model.

As stated in the introduction, the development of the AMAU descriptive system was based on a theoretical model of HRQoL. Items were generally based on hypothesised elements of this model, and were tested on hospital and community subjects. The responses to each item were subjected to factor analysis to detect underlying themes (factors) which suggest unique dimensions of HRQoL. This section examines the item content of each factor derived from the analysis and compares each resulting scale with the hypothesised structure of HRQoL (Table 1).

Illness was hypothesised to consist of pain, reliance on medical aids and receiving medical treatment. The analysis showed that this construct was supported with three items from two of the hypothesised elements.

Independent living was hypothesised to consist of bodily care, activities of daily living (ADL), and safety and security. Bodily care and ADL were covered in the scale. An item referring to mobility (from the hypothesised 'physical ability' dimension) loaded in this factor: the item referred to the ability to get around the local community. This was consistent with the theme of this dimension and is essential for independent living.

Physical ability referred to physical performance, sensory functions, mobility and fatigue, and sexual relationships. This hypothesised dimension of HRQoL did not appear in the empirical model. Physical ability elements loaded on 'independent living' as mobility (getting around the community), in 'social relationships' as satisfaction with sex, and in 'psychological well-being'

as rest/fatigue (sleep).   A separate factor for *physical senses* was however identified during the analysis, which included sight and hearing.

Psychological state referred to anxiety/depression, cognitive ability, emotional fulfilment and memory. The analysis showed that this construct was supported with only one item from the hypothesised elements: anxiety/depression. Other items which loaded on this factor included pain (from the 'illness' dimension) and sleep (from the 'physical ability' dimension). While not consistent with the hypothesised model, these findings were consistent with clinical and psychological theories of pain and sleep.

Social relationships was hypothesised to consist of communication, family role, intimacy, social function and work satisfaction. The analysis showed that this construct was supported with a social relationships item and a family role item. An item from the hypothesised 'physical ability' dimension, satisfaction with sex, loaded on this factor. Undoubtedly this aspect of life reflects a very personal and intimate aspect of social relationships.

These analyses indicate that the derived model of HRQoL differs slightly from the original hypothesised model, although it is consistent with it. The derived model was logically coherent, covered the principal dimensions of HRQoL and possessed good psychometric properties.

3.4.1   Factor structure and correlation between scales

MAU theory requires that the different scales within an instrument must be orthogonal to avoid double counting and to demonstrate independence (Winterfeldt & Edwards 1986). Construction of a valid descriptive system therefore requires orthogonality. This is an extreme form of discriminate validity—which is where scores on one scale are independent of scores obtained on another scale. Within the conventional bounds for psychometric instrument construction (Anastasi 1976) it is accepted that correlations below 0.60 indicate discriminate validity.

In the AMAU the average correlation between scales is 0.31, and the range is from 0.44 to 0.21. From this it may be concluded that the scales are orthogonal (determined through varimax rotation; Table 7) and possess discriminate validity (Table 9). It would appear that the scales reliably and independently measure different dimensions of HRQoL.

In addition to structural orthogonality, MAU theory requires preference independence: the preference score for one item should not depend upon the level of another item. This property is not generally tested but is inferred from the logical relationship between the items (Feeny,

Furlong et al. 1995). The present survey did not permit the testing of this property, although it is being explored in work-in-progress.

Table 9 Correlation matrix of the scales (Pearson's correlation coefficients)

	Illness	Independent living	Social relationships	Psychological well-being
Independent living	.44			
Social relationships	.31	.35		
Psychological well-being	.36	.32	.38	
Physical senses	.20	.22	.21	.28

3.4.2 Additional Items

The items appearing in the above instrument provided the most parsimonious and logical model. However, during the factor analyses—see stage 3 above—it was noted that several items could be replaced by alternative items which had almost identical inclusion criteria and content. Given that the next stage in the development of the AMAU is scaling, the items that "did not quite make it" will be included. It is possible that some items currently included may perform poorly in the TTO exercises or prove to contribute little to the overall disutility score within a factor. After scaling, the item importance will be tested. Therefore having additional and alternative equivalent items is advantageous.

It is therefore intended that for each scale an additional item will be incorporated during the TTO scaling interviews. In each case this additional item will be selected so that it meets the inclusion criteria outlined above in Section 3.3.2.

3.5 Preliminary performance indicators of the AMAU

A 'useful' MAU instrument is required to be sensitive enough to detect differences between sub-populations. As described above in Section 2.3.1 the sample was drawn in a manner which allowed differences between hospital and community subjects to be examined.

The ability of the final instrument was initially evaluated through analysis of the sample sub-groups; by hospital/community group, age, sex, socioeconomic status (SES) and marital status. It is acknowledged that legitimate evaluation of the sensitivity of an instrument is required with a different sample of subjects; nonetheless, examination of differences within the original sample provide some indicators of the instrument's potential.

The QoL scores of each of the five factors, and the total QoL score of the AMAU, using a simple additive unweighted model, were evaluated for the ability to detect differences between various sub-populations among subjects. This was undertaken by summing the scores of each item within each factor, and for the total score, all five factors were summed.

In the following description of sub-populations a *significant difference* refers to a statistically significant difference ($p < 0.05$) using the Mann-Whitney U-test for non-parametric data unless otherwise stated.

3.5.1 Hospital, community and adolescent subjects

Most of the scales of the AMAU were found to distinguish between the adult (hospital and community) and the adolescent samples (Figure 3). In general the adolescents had significantly higher QoL, which was expected as adolescents were a selected population (students well enough to be at school), younger and actively participating in a school environment.

Table 10   Analysis of differences between scales' scores in hospital, community and adolescent subjects.

| | Group Differences[†] | | | Total[‡] |
	Hospital vs Community	Hospital vs Adolescents	Community vs Adolescents	Overall ANOVA
Illness	+	+	+	+
Independent living	+	+	ns	+
Social relationships	+	+	ns	+
Psychological well-being	+	+	+	+
Physical senses	ns	+	+	+
Total QoL score	+	+	+	+

† Mann-Whitney U test

‡ One-way Analysis of Variance (ANOVA)

+ p < 0.05

ns p > 0.05

Figure 3 Scale scores for three groups; hospital cases, community cases and adolescents (NOTE: a high score indicates lower Quality of Life).
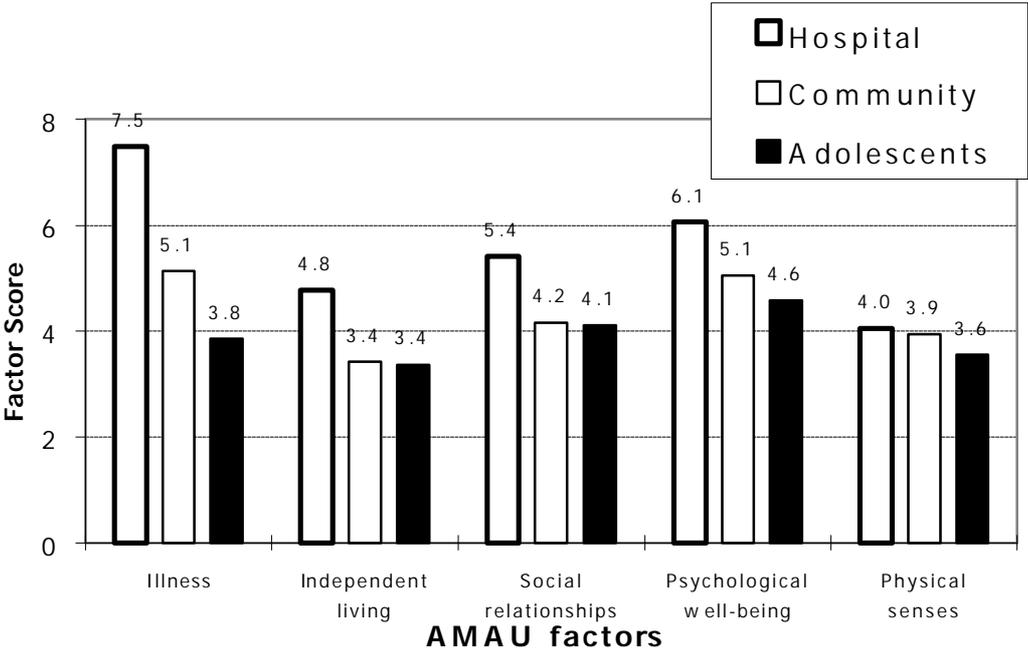
The scale scores significantly distinguished hospital patients from adolescents on all scales (Table 10). Patients reported significantly higher scores (ie worse QoL) than community cases on all scales except the Physical senses scale. Community case scores were significantly higher than adolescents' on the Illness, Psychological well-being and Physical senses scales.

These results were consistent with expectations. Hospital cases have poorer health than people in the community and hence they scored lower HRQoL. Adolescents attending school were expected to have less health related problems than either patients or community cases. This trend was reflected in the scale scores and in the total summed instrument scores.

3.5.2 Age groups

It was expected as the age of subjects increased, HRQoL would decrease for community cases, but not for patients affected by illness. The data supported this expectation as shown in Figures 4 & 5. Figure 4 shows a scatter plot of age vs HRQoL for the community sample and Figure 5 for the hospital patients. When all cases were pooled and analysed it was found that age was significantly correlated with all scales and total HRQoL, as shown in Table 11.

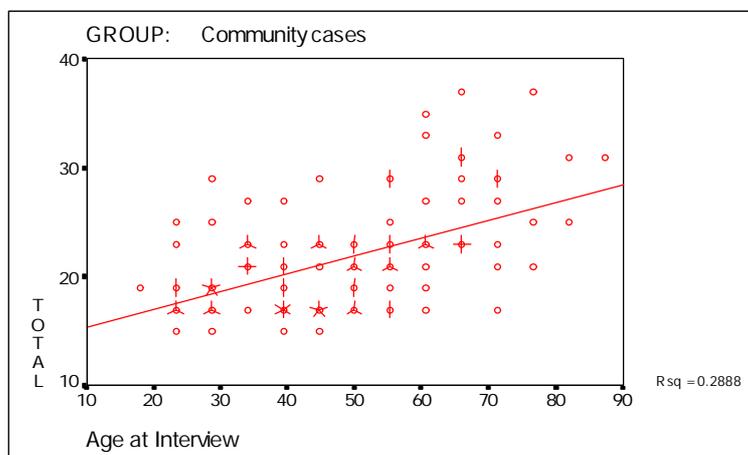Figure 4 Scatter plot of total HRQoL score vs age of community cases

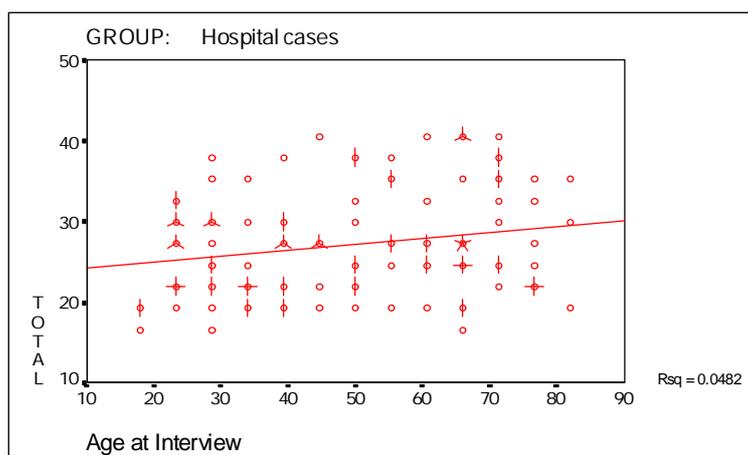Figure 5 Scatter plot of total HRQoL score vs age of hospital patients

Table 11   Relationship between age of cases and scale scores (Spearman's *r*)

	Hospital	Community	Adolescent	All
Illness	.18*	.60**	ns	.52**
Independent living	ns	.53**	ns	.32**
Social relationships	ns	ns	ns	.15**
Psychological well-being	ns	.21*	ns	.29**
Physical senses	.25**	.32**	ns	.28**
Total QoL score	.20*	.51**	ns	.49**

Note: * p < 0.05, ** p < 0.01

With respect to the adolescent sample, age was not correlated with HRQoL scores (Table 11). This was expected as the sample were all from Year 10 and therefore formed a single age cohort.

Social relationships and Psychological well-being were weakly correlated with age suggesting that these factors are not age dependent. As expected, QoL related to Illness and Physical senses declined with increasing age.

3.5.3   Gender

When the data for each scale were examined, males scored higher (lower QoL) than females on the Social relations scale, and lower (higher QoL) on the Psychological well-being scale when compared with females (MANOVA, F=4.70, p=0.03, and F=5.25, p=0.02 respectively). There was, however, no overall significant difference between scores.

In the adolescent group, males scored lower (higher QoL) than females on the Illness scale (Kruskall-Wallis, $\square^2 = 6.53$, p=0.01) and the Independent living scale (Kruskall-Wallis, $\square^2 = 3.78$, p = 0.05).

3.5.4   Educational level

The educational level achieved by subjects was divided into 5 categories: no education, primary only, completed secondary school, trade or technical college, and completed tertiary (Figure 6). Total HRQoL tended to decrease with decreasing educational level (ANOVA, df=209, F=3.36, p=0.02). This trend was also significant in the Illness, Independent living and Psychological well-being scales (Spearman's r < 0.05).

Figure 6 HRQoL of hospital patients and community cases by educational level.
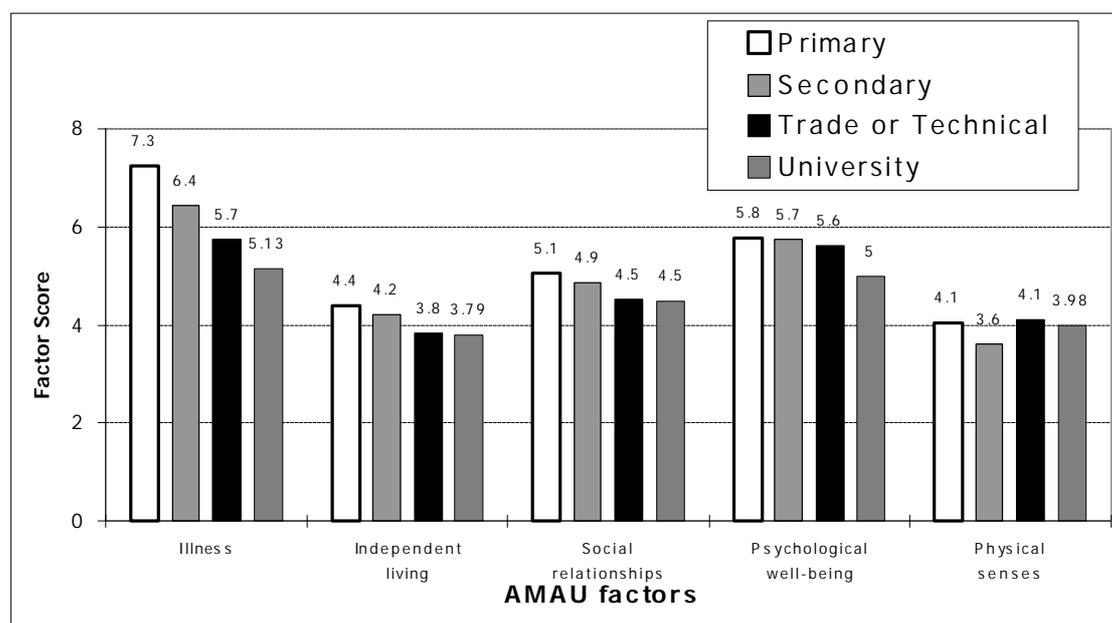
3.5.5 Partnered status

The AMAU scales' scores and the total score did not significantly distinguish between cases who had a partner (married/defacto) and those subjects who were unpartnered (single, widowed, divorced). However, there was some evidence unpartnered subjects tended to score lower on the Illness scale ($p = 0.059$), Independent living scale ($p = 0.089$), and the Psychological well-being scale ($p = 0.053$). The data are given in Figure 7.

Figure 7 HRQoL of hospital patients and community cases by marital status (partnered or unpartnered)
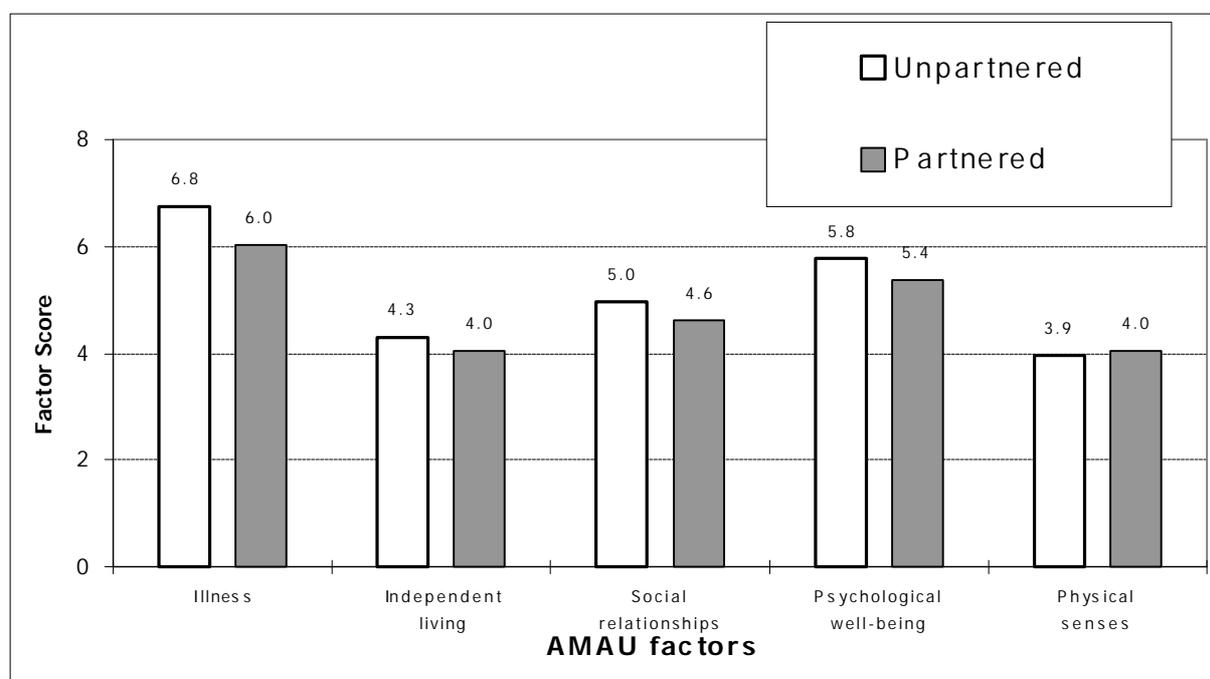
4   Discussion

This paper has described the further development of the AMAU, building on the work of Hawthorne and Richardson (1996); specifically the construction of the AMAU descriptive system for measuring HRQoL. This system consists of five dimensions or scales with three items in each dimension, a total of 15 items. The structure outlined in the paper is consistent with the hypothesised structure and dimensions of HRQoL defined earlier in the project.

The psychometric properties of the instrument are reported, including preliminary indicators of its capabilities. The results suggest the AMAU has the ability to distinguish levels of health evident in hospital patients, community cases and adolescents, between age groups, gender and educational level. In addition, interviews with subjects show it is very easy to understand and complete with minimal difficulty. The AMAU is thus a short and convenient measure of HRQoL.

4.1   Domain-referenced validity: the factor structure and the theoretical concept of HRQoL

It was hypothesised that HRQoL comprised five dimensions: illness, independent living, physical ability, psychological state and social interaction, each with a number of elements (Table 1).

The factorial structure of the final instrument suggests this hypothesised model was broadly supported (Table 7), although the identified factors varied slightly from the original model. These variations revolved around inclusion on factors of items originally hypothesised as belonging to other HRQoL dimensions. As explained in the text, in all cases these were logically consistent with the item content and clinical theory (Section 3.4).

The only hypothesised dimension which was not supported was the Physical ability dimension. The data suggested the original content of this dimension was too disparate and shared commonality with several other dimensions. The factor ultimately constructed representing this dimension resulted is a more elegant scale centred around the key physical senses: sight, hearing and communication ability.

Overall, subject to this caveat, the hypothesised structure was supported. As indicated by the match between the theoretical model and the empirically derived model suggests the AMAU possesses domain-referenced validity.

4.2   Issues in "ceiling" and "floor" effects of self-report Likert/Guttman scales

Foddy (1993) pointed out that during psychometric measurement, subjects were likely to agree with a statement rather than disagree or select an uncomfortable (for them) response. When applied to HRQoL measurement this suggests there may be 'ceiling' or 'floor' effects. These effects confound measurement by not allowing the full breadth of health states to be measured. There are two ways these could arise: through poor item sensitivity or through sample bias.

In the construction of the AMAU steps were taken to avoid poor item sensitivity; both through the iterative item development stage outlined in Hawthorne and Richardson (1996) and through the item selection procedures described in this paper. These procedures were designed to result in the selection of items reflecting real health states that are actually experienced by people with ill health. The consequence is that the lower boundaries (poor health) reflect real restrictions rather than conceptualised worst health states which are 'unreal' and which therefore may lead to inflated estimates of intervention efficacy.

With respect to sample bias, where HRQoL measurement samples are drawn from hospital patients only, the sample may be skewed towards the unhealthy end of the health spectrum. The sample used for instrument construction in this project explicitly sought to minimise this bias through selection of both patients and community cases, with the attendant problem of a 'ceiling' effect: healthy people will report no or little loss of QoL.

This, however, is the normal situation for QoL instruments: a meta-analysis of 16 studies containing data on life satisfaction reported the mean scores, when computed as '% of scale maximum' (%SM) showed all the studies' means were within 70–80%SM (Cummins 1995).

The value obtained for the AMAU was at the high end of this range, being 78.9%SM for all cases. When disaggregated by hospital and community cohorts the respective values were 73.0%SM and 85.4%SM. Given the method of drawing the community sample—through letter

and telephone follow-up—the skew towards the healthy end of the spectrum was expected. It does suggest, though, that the results were broadly consistent with those obtained elsewhere.

4.3 Differences between the AMAU and the preliminary Australian MAU instrument

The preliminary analysis reported previously (Hawthorne and Richardson 1996) was based on analysis of 106 hospital patients. There were no community cases. It is likely that this difference in the construction sample explains the slight differences between the preliminary instrument and the AMAU reported in this paper.

The two instruments are, however, remarkably similar. In both versions there were five scales, each of three items. The psychometric properties of the two instruments were virtually identical; the preliminary instrument explaining 69% of variance compared with the AMAU's 65%, the eigenvalues for each of the scales closely matched each other, in the preliminary instrument none of the items cross-loaded on varimax rotation whereas three items did in the AMAU, the reliability analyses suggested the two versions were equally reliable (α=0.78 and 0.80 for the preliminary and AMAU versions respectively), and the percentage scale scores for hospital patients only were 72%SM for the preliminary version and 73%SM for the AMAU.

In fact the only significant differences related to three items; those covering pain, sex and hearing. Neither of the first two items were included in the preliminary instrument although they were suggested as additional items which could have been included. They were included in the AMAU, at the expense of items covering purpose in life (Scale 3: Social relationships) and cognition (Scale 4: Psychological state). In both cases the inclusion of these items makes for a more coherent scale with respect to scale concept and interpretation.

The third item, hearing, replaced an item covering bodily senses. This item referred to the physical senses, including vision and hearing. Since vision was covered by another item it could be argued that double-counting was occurring in this scale. On logical grounds, then, the inclusion of an item probing hearing alone suggests a 'cleaner' solution.

That the AMAU is a marginally 'dirtier' solution than the preliminary instrument—in the sense that three items cross-loaded as described—may be attributed to inclusion of the community sample; while enlarging the sample size this also had the effect of increasing the variance within the data (as explained above the two groups were systematically different). This should

be placed in the broader context: that both versions contain the same internal structure—with five orthogonal factors or scales—strongly suggests the hypothesised model has indeed tapped into and adequately represented HRQoL.

That two separate samples have shown the same factorial structure, with virtually identical psychometric properties provides preliminary evidence for the validity and reliability of the instrument.

4.4   Comparison with other instruments

A small number of utility instruments are available for use in evaluations involving cost utility analysis. The major instruments are the EUROQoL, 15D, Health Utility Index (HUI, Versions I, II & III), QWB and Rosser/Kind. Each of these were constructed differently, have different properties and each would be preferred in a variety of research applications. The development of the AMAU was a response to flaws in these instruments. Our assessment of the strengths and weaknesses of these instruments is summarised in Table 12. Judgement of the AMAU is made on the basis of the evidence presented in this report.

Table 12:   Evaluation of MAU instruments

	Instrument							
	EUROQoL	HUI			15D	QWB	Rosser - Kind	AMAU
		I	II	III				
Country	UK	Canada			Finland	USA	UK	Australia
Focus	Health status	Health status	Health status	Bodily function	Function status	Function status	Morbidity	Health status
N. dimensions	5	4	7	3	15	3	2	5
N. items	5	4	7	24	15	18	2	15
Scaling technique	TTO	RS/TTO	RS/SG	RS/SG	RS	RS	ME	RS/TTO
Health profile	⌉	⌉	⌉	⌉	⌉	⌋	⌋	⌉
Single indice	⌉	⌉	⌉	⌉	⌉	⌉	⌉	⌉
Sensitivity	⌋	⌋	?	⌉	⌉	?	⌋	⌉
Psychometric structure	⌋				⌉		⌋	⌉
Reliability	⌉		⌉		⌉	⌉		⌉*
Validity—psychometric	?	?			⌉	⌉	⌋	*
Validity—economic	?	⌉	⌉	⌉	⌋	⌋	⌋	⌉*
Ease of use	⌉	⌉	⌉	⌉	⌉	?	⌉	⌉

Notes: ME = magnitude estimation

RS = rating scale

TTO = time trade-off

* = work currently in progress. The reliability of the AMAU is based on internal consistency estimates only

References: Drummond, Stoddart et al. 1987; Rosser 1993; Sintonen 1994; Feeny, Furlong et al. 1995; MVHGroup 1995; Sintonen 1995; Torrance, Furlong et al. 1995; Williams 1995; Feeny, Torrance et al. 1996; Kind 1996.

Summary and future work

The AMAU has been developed from a strong theoretical basis using sophisticated psychometric construction techniques. From the literature this does not appear to have occurred with other instruments. The factorial structure and reliability estimates all provide evidence of its sound internal coherence and construct validity.

The findings also indicate that the instrument is suitable for scaling and validation as a QoL/QALY instrument; this is currently being undertaken. In addition, the preliminary evidence relating to its ability to differentiate between the health states of construction sample sub-groups and a separate sample of adolescents suggests the instrument possesses sensitivity and discriminate validity.

In short, as demonstrated in this paper, the descriptive system of the AMAU has sound psychometric properties and is ready for use in validation and evaluation studies as an unweighted HRQoL measure.

*

As noted earlier the selection of the appropriate model for combining item scores is problematical and has received little attention in the empirical literature. A difficulty that does not appear to have been satisfactorily overcome is that non-orthogonal items have been combined in such a way that the negative effect of one aspect of health has been separately included through two or more items. This results in the 'double counting' of this negative effect.

In the present project various models will be tested empirically. However, the underlying approach to the above problem of structural correlation is as follows. Each of the five dimensions will be independently modelled, using either an additive or multiplicative method. The utility value of the all-worst health state in each dimension will be independently determined and the contribution of that dimension to the total disutility constrained by this value. As the greatest correlation occurs between items within dimensions this prevents 'double-counting' within dimensions affecting the overall score. In this way, great sensitivity within dimensions can be combined with score validity. As presented in this paper the AMAU has achieved orthogonality between dimensions. These may therefore be combined without double-counting. The choice of an additive or multiplicative model will be determined empirically.

The final stage of the project will be to carry out a conventional process of validation. In part this requires a demonstration of construct validity. Validation will include testing the AMAU against other validated instruments to determine whether or not it is an acceptable HRQoL instrument.

*

In sum, the results reported above indicate that a wide range of health states may be described by the AMAU's limited number of dimensions and that these dimensions broadly correspond with those found in the literature. The orthogonality of the dimensions supports the assertion that it is a valid descriptive system which may form the basis for a reliable, valid and sensitive MAU instrument.

REFERENCES

Anastasi, A. (1976). Psychological Testing. New York, Macmillan Publishing.

Burrows, C. and K. Brown (1993). "QALYs for resource allocation: probably not and certainly not now." Australian Journal of Public Health **17**(3): 278-286.

Cook, J. and J. Richardson (1993). Quality of life measurement of patients receiving treatments for gallstone disease: options, issues and results. Working Paper No. 33. Melbourne, National Centre for Health Program Evaluation.

Cummins, R. (1995). The comprehensive quality of life scale: theory and development. Proceedings Health Outcomes and Quality of Life Measurement Conference, Canberra, Australian Institute of Health and Welfare.

Drummond, M. and L. Davies (1991). "Economic analysis alongside clinical trials: revisiting the methodological issues." Int. J. of Technology Assessment in Health Care **7**(4): 561-573.

Drummond, M., G. Stoddart, et al. (1987). Methods for the Economic Evaluation of Health Care Programmes. Oxford, Oxford University Press.

Feeny, D., W. Furlong, et al. (1995). "Multi-attribute health status classification systems." PharmacoEconomics **7**(6): 489-502.

Feeny, D., G. Torrance, et al. (1996). Health utilities index. Quality of Life and Pharmacoeconomics in clinical trials. S. B. Philadelphia, Lippincott-Raven Publishers.

Foddy, W. (1993). Constructing Questions for Interviews and Questionnaires: Theory and Practice in Social Research. Cambridge, Cambridge University Press.

Hawthorne, G. and J. Richardson (1996). An Australian MAU/QALY Instrument: Rationale and Preliminary Results. Working Paper 49. Melbourne, Centre for Health Program Evaluation.

Kind, P. (1996). The EUROQoL instrument: and index of health-related quality of life. Quality of Life and Pharmacoeconomics in clinical trials. B. Spilker. Philadelpha, Lippincott-Raven Publishers.

MVHGroup (1995). The Measurement and Valuation of Health: Final Report on the Modelling of Valuation Tariffs. York, Centre for Health Economics.

Nord, E., J. Richardson, et al. (1993). "Social evaluation of health care versus personal evaluation of health states." International J. Tech. Assessment in Health Care **9**(4): 463-478.

Norusis, M. (1993). SPSS for Windows. Chicago, SPSS Inc.

Richardson, J. (1991). "Economic assessment of health care: theory and practice." The Australian Economic Review **1**(4-19).

Rosser, R. (1993). A health index and output measure. Quality of Life Assessment: Key Issues in the 1990s. S. Walker and R. Rosser. Dordrecht, Kluwer Academic Publishers.

Sintonen, H. (1994). The 15D measure of health-related quality of life: reliability, validity and sensitivity of its health state descriptive system. Working Paper 41. Melbourne, National Centre for Health Program Evaluation.

Sintonen, H. (1995). The 15D measure of health-related quality of life: feasibility, reliability and validity of its valuation system. Working Paper 42. Melbourne, National Centre for Health Program Evaluation.

Torrance, G., W. Furlong, et al. (1995). "Multi-attribute preference functions: health utilities index." PharmacoEconomics **7**(6): 503-520.

Ware, J. E. (1987). "Standards for validating health measures: definition and content." J Chron Dis **40**(6): 473-480.

WHO (1958). The First Ten Years of the World Health Organization. Geneva, World Health Organization.

Williams, A. (1995). The Measurement and Valuation of Health: A Chronicle. York, Centre for Health Economics.

APPENDIX 1: THE AMAU INSTRUMENT

1   ILLNESS

1.1   Concerning my use of prescribed medicines:
- A.   I do not or rarely use any medicines at all.
- B.   I use one or two medicinal drugs regularly.
- C.   I need to use three or four medicinal drugs regularly.
- D.   I use five or more medicinal drugs regularly.

1.2   To what extent do I rely on medicines or a medical aid? (NOT glasses or a hearing aid.) (Eg. walking frame, wheelchair, prosthesis etc.)
- A.   I do not use any medicines and/or medical aids.
- B.   I occasionally use medicines and/or medical aids.
- C.   I regularly use medicines and/or medical aids.
- D.   I have to constantly take medicines or use a medical aid.

1.3   Do I require regular medical treatment from a doctor or other health professional?
- A.   I do not need regular medical treatment.
- B.   Although I have some regular medical treatment, I am not dependent on this.
- C.   I am dependent on having regular medical treatment.
- D.   My life is dependent upon regular medical treatment.

2   INDEPENDENT LIVING

2.1   Do I need any help looking after myself?
- A.   I need no help at all.
- B.   Occasionally I need some help with looking after myself.
- C.   I need help with the more difficult personal care tasks.
- D.   I need daily help with most or all personal care tasks.

2.2   Thinking about doing household tasks (eg. preparing food, gardening, using the video recorder, radio, CD player, telephone or washing the car):
- A.   I have no difficulty doing these sort of things.
- B.   I find some of these things difficult, but I manage to do them on my own.
- C.   Many of these things are difficult, and I need help to do them.
- D.   I cannot do these things by myself, and I cannot get help to do them.

2.3 **Thinking about how easily I can get around:**
- A. I get around my home and community by myself without any difficulty.
- B. I find it difficult to get around my home and community by myself.
- C. I cannot get around the community by myself, but I can get around my home with some difficulty.
- D. I cannot get around either the community or my home by myself.

3   SOCIAL RELATIONSHIPS

3.1 **Because of my health, my relationships (eg. with my friends, partner or parents) generally are:**
- A. Are very close and warm.
- B. Are sometimes close and warm.
- C. Are seldom close and warm.
- D. I have no close and warm relationships.

3.2 **To what extent am I satisfied with sex in my life (even if I don't engage in sex):**
- A. I have no problems with my sexual activities.
- B. I am satisfied overall.
- C. I have some difficulties.
- D. My sex life is difficult for me.
- E. I am completely dissatisfied with sex in my life.

3.3 **Thinking about my health and my relationship with my family:**
- A. My role in the family is unaffected by my health.
- B. There are some parts of my family role I cannot carry out.
- C. There are many parts of my family role I cannot carry out.
- D. I feel I have lost my role in the family.

4   PHYSICAL SENSES

4.1 **Thinking about my vision, including when using my glasses or contact lenses if needed:**
- A. I see normally.
- B. I have some difficulty focusing on things, or I do not see them sharply. Eg. small print, a newspaper, or seeing objects in the distance.
- C. I have a lot of difficulty seeing things. My vision is blurred. Eg. I can see just enough to get by with.
- D. I only see general shapes, or am blind. Eg. I need a guide to move around.

4.2   Thinking about my hearing, including using my hearing aid if needed:
   A.   I hear normally.
   B.   I have some difficulty hearing or I do not hear clearly. Eg. I ask people to speak up, or turn up the TV or radio volume.
   C.   I have difficulty hearing things clearly. Often I do not understand what is said. Eg. I usually do not take part in conversations because I cannot hear what is said.
   D.   I hear very little indeed. Eg. I cannot fully understand loud voices speaking directly to me.

4.3   When I communicate with others (eg. by talking, listening, writing or signing):
   A.   I have no trouble speaking to them or understanding what they are saying.
   B.   I have some difficulty being understood by people who do not know me. I have no trouble understanding what others are saying to me.
   C.   I am only understood by people who know me well. I have great trouble understanding what others are saying to me.
   D.   I cannot adequately communicate with others.

5   PSYCHOLOGICAL WELL-BEING

5.1   If I think about how I sleep:
   A.   I am able to sleep without difficulty most of the time.
   B.   My sleep is interrupted some of the time, but I am usually able to go back to sleep without difficulty.
   C.   My sleep is interrupted most nights, but I am usually able to go back to sleep without difficulty.
   D.   I sleep in short bursts only. I am awake most of the night.

5.2   Thinking about how I generally feel:
   A.   I do not feel anxious, worried or depressed.
   B.   I am slightly anxious, worried or depressed.
   C.   I feel moderately anxious, worried or depressed.
   D.   I am extremely anxious and depressed.

5.3   Does pain interfere with my life at all?
   A.   No, it does not interfere.
   B.   Yes, it interferes slightly.
   C.   Yes, it interferes a lot.
   D.   Yes, it constantly interferes.