Comparison of the measurement properties of the AQoL and SF-6D in critical illness

Elizabeth H Skinner, Linda Denehy, Stephen Warrillow and Graeme Hawthorne

ABSTRACT

Objective: Multiattribute utility (MAU) instruments are short instruments measuring quality of life, health utility scores and treatment cost-effectiveness. Many studies have compared MAU instruments, but few have involved intensive care unit patients. Our aim was to compare the measurement properties of two MAUs, the assessment of quality of life (AQoL) and Medical Outcomes Study Short Form 6D (SF-6D), in a sample of patients with critical illness.

Methods: Prospective observational study conducted in an 18-bed mixed tertiary Australian ICU. Eligibility criteria were: admitted to the ICU > 48 hours, aged > 18 years, and not imminently at risk of death. Participants completed the AQoL and SF-6D on admission to the ICU as a "then-test" of pre-ICU status, and 6 months after ICU discharge. We assessed the reliability, validity, sensitivity and responsiveness of the instruments.

Results: Median age was 61 years (interquartile range [IQR], 49–73 years)], 60% were men, and the median Acute Physiology and Chronic Health Evaluation II score was 17 (IQR, 13–21). Cronbach's α was acceptable for the AQoL ($\alpha = 0.81$) but not for the SF-6D ($\alpha = 0.65$). The AQoL and SF-6D showed evidence of validity but, despite moderate agreement between their utilities, their scores were not interchangeable. This was likely due to the SF-6D's truncated scoring range. The AQoL was predictive of hospital readmission but the responsiveness and sensitivity of the instruments varied by clinical condition.

Conclusions: The AQoL and SF-6D demonstrated acceptable measurement properties in the ICU population, but the findings raised questions about the reliability and predictive power of the SF-6D. Further research is required to determine the most appropriate instrument for use in measuring health utility in the ICU population.

Crit Care Resusc 2013; 15: 205–212

Measurement of health utility has traditionally been used in cost–utility analysis (CUA), and CUA is considered to be the gold standard for economic evaluation.[1] Clinicians perceive that measuring health utilities as part of clinical care is clinically relevant and of benefit to patient–clinician assessment, relationships, communication and patient management.[2] Health utility scores are calculated using purpose-designed multiattribute utility (MAU) instruments, such as the assessment of quality of life (AQoL),[3] health utilities index (HUI),[4] EuroQoL 5D (EQ-5D)[5] or Medical Outcomes Study Short Form 6D (SF-6D) instruments.[6] MAU instruments are shorter and easier to use than other generic measures, a characteristic which is particularly useful in the intensive care unit. MAU instruments allow the weighting of patient health states using previously described preferences for particular health states, and utilities can be used to assess clinical outcomes and treatment cost-effectiveness.

Measurement of the health utility of patient premorbid health states and health-related quality of life (HRQoL) on admission to the ICU may be useful in clinical decision making and the prediction of clinical outcomes.[7] Understanding premorbid health-related quality of life is essential to understanding the illness and recovery trajectories after ICU discharge.[8] Use of MAU instruments in the clinical setting may become increasingly important in the ICU for resource allocation decisions at the patient level, in addition to the economic evaluation of intensive care treatment, which is already subject to considerable attention[9] in the context of burgeoning ICU costs.[10] However, accurate measurement of utilities relies on robust instrument properties such as reliability, validity and responsiveness. Direct head-to-head comparisons of MAU instruments have consistently shown that psychometric differences lead to differences in interpretation of measured change in clinical outcomes and estimates of treatment cost-effectiveness.[11-13] These differences may affect clinical decision making, resource allocation and health policy.[9,14,15] Few reports have examined the reliability and validity of MAU instruments in people with critical illness,[14] and only one report has compared measurement properties in this population.[16]

The objective of our study was to compare the measurement properties (reliability, validity, sensitivity, responsiveness and predictive power) of two MAU instruments, the AQoL and SF-6D. The secondary aim was to report health utilities and quality-adjusted life years (QALYs) in an Australian population sample with critical illness.

Methods

Setting and participants

Our prospective cohort study was conducted in an Australian general 18-bed tertiary ICU during October 2006 to March 2007. The inclusion criteria were age ⩾ 18 years, ICU length of stay > 48 hours, and the patient or an immediate family member having adequate English skills to complete the questionnaires, verbally or in writing. Exclusion criteria were likely imminent death within 48 hours of ICU admission, admitted with new neurological insult (eg, intracerebral haemorrhage or spinal cord injury) or an acute psychiatric disorder. Further key methodological details are described elsewhere.[17] The institutional human research ethics committee approved the study and informed consent was obtained.

Procedure

Participants completed HRQoL questionnaires as soon as they were able after ICU admission. This was completed as a "then-test" (ie, assessment of their status before the health crisis leading to admission to the ICU) and at 6 months after ICU discharge. The baseline questionnaires were completed during face-to-face interviews unless participants preferred to complete them alone. Follow-up questionnaires were completed by telephone interview unless a mailed copy was requested. When participants were unable to be contacted by telephone, five further attempts were made.

Measures

Demographic information was collected as it was in the index cohort study.[17] Participants were asked at both time points whether they were living at home, in paid employment, or receiving any organised community services (eg, council home help). Their discharge destination was recorded, ie, home with no additional support, home with extra support (eg, in-home physiotherapy), inpatient rehabilitation, or other (eg, interhospital transfer). Data that we were unable to obtain from participants were extracted from the medical record, where available.

General health was assessed with the 36-item Medical Outcomes Study Short Form 36 Version 1 (SF-36).[18] For the purpose of this study, the response categories "excellent", "very good" and "good" on the general health status item (question 1) were collapsed into "good", and the response categories "fair" and "poor" were collapsed into "poor". To assess health utility, the SF-6D[19] and AQoL[3] were completed.

Table 1. Demographic data and severity of illness of study participants (*n* = 67)

Variable	No (%)*
Mean age, years (SD)	59.8 (14.5)
Male, %	60
Median APACHE II score (IQR)	17 (13–21)
Diagnosis	
Cardiac; cardiological (combined %)	16; 5 (31)†
General medical; sepsis (combined %)	13; 8 (31)
General surgical (%)	7 (10)
Respiratory (%)	6 (9)
Thoracic, vascular or orthopaedic surgical (%)	9 (13)
Liver transplant (%)	3 (4)
Medical procedure (%)	30 (45)
Surgical procedure (%)	37 (55)
MV required (%)	48 (72)
Median time on MV, hours (IQR)	42 (16–87)
Median time in ICU, hours (IQR)	101 (68–149)
Readmitted to ICU (%)	12 (18)
Length of stay in ICU ⩾ 5 days (%)	33 (49)
Median length of stay in hospital, days (IQR)	17 (11–32)

APACHE = acute physiology and chronic health evaluation. IQR = interquartile range. MV = mechanical ventilation. ICU = intensive care unit. * Unless otherwise stated. † 24% were for open heart surgery.

The AQoL instrument

The AQoL instrument is an MAU which assesses handicap arising from health states.[3] It performs well in comparison to other MAU instruments,[20] although it has not been validated in the critical care setting. The descriptive system consists of 15 items organised into five dimensions (illness, independent living, social relationships, physical senses and psychological wellbeing). The preference weights were obtained from a representative sample of the Australian population using the time trade-off (TTO) method.[21] Utility scores cover the spectrum from 1.00 (best HRQoL state) to –0.04 (worst HRQoL state) where 0.00 is a death-equivalent state.

The SF-6D instrument

Based on the descriptive system of the SF-36,[18] the SF-6D was developed for use in situations where the SF-36 had been administered and an economic evaluation was desired.[19] The reason for selection of the SF-6D in this study is that the SF-36 is recommended for use in the ICU population.[22] If it performed as well as a specialist MAU instrument, it could be recommended, thus obviating the need for including another MAU instrument for economic evaluation. The SF-6D uses 11 items from the SF-36, combined to categorise patients in six domains (physical functioning, role limitations, social functioning, pain, mental health and vitality), and scored using utility weights obtained from a representative sample of the United Kingdom population, using the standard gamble method.[19] Utility scores range from 1.00 (best HRQoL state) to 0.30 (worst HRQoL state).[19] Although the SF-6D has been directly compared against other MAU instruments, such as the EQ-

Table 2. Validity and sensitivity of the AQoL and SF-6D

Variable	Status (*n*)	AQoL*		SF-6D	
		Mean	SD	Mean	SD
Before ICU admission ($n=89$)†					
Health status	Good (31)	0.64	0.29	0.64	0.13
	Poor (36)	0.35	0.28	0.51	0.12
	Statistics‡	$t=4.27$***	–	$t=4.41$***	–
	RE	1.00	–	1.07	–
Sex	Male (40)	0.53	0.34	0.58	0.15
	Female (27)	0.43	0.28	0.57	0.13
	Statistics‡	$t=1.35$	–	$t=0.21$	–
	RE	NA	–	–	–
Age group (years)	27–55 (23)	0.43	0.32	0.56	0.14
	56–70 (26)	0.49	0.32	0.58	0.15
	71–82 (18)	0.55	0.32	0.58	0.14
	Statistics§	F = 0.74	–	F = 0.25	–
	RE	NA	–	–	–
ICU outcomes and utilities 6 months after admission					
Rehabilitation	Yes (31)	0.66	0.33	0.69	0.15
	No (36)	0.43	0.29	0.58	0.11
	Statistics‡	$t=-2.98$***	–	$t=-3.39$***	–
	RE	1.00	–	1.29	–
Home discharge¶	Home, no support (24)	0.60	0.33	0.63	0.13
	Home, supported (20)	0.45	0.31	0.63	0.13
	Statistics‡	$t=1.57$	–	$t=0.05$	–
	RE	N/A	–	–	–
Increased requirement for community services††	Yes (10)	0.32	0.29	0.59	0.14
	No (50)	0.60	0.32	0.64	0.14
	Statistics‡	$t=2.33$**	–	$t=1.03$	–
	RE	1.00	–	0.20	–
6-month general health status‡‡	Good (36)	0.74	0.21	0.71	0.11
	Poor (31)	0.29	0.27	0.54	0.12
	Statistics‡	$t=7.68$***	–	$t=5.96$***	–
	RE	1.00	–	0.60	–

AQoL = Assessment of Quality of Life. SF-6D = Short Form 6D. ICU = intensive care unit. RE = relative efficiency. NA = not applicable. * AQoL was set as the denominator for all tests, for ease of understanding. † Construct validity calculated on baseline data using preadmission AQoL and SF-6D utilities, sex and age. ‡ Independent *t* test, df = 65. § Analysis of variance, df = 3,85. ¶ Excludes those not discharged from hospital or residential care. †† Data missing from 10 cases. ‡‡ Criterion validity (calculated on follow-up utility scores), health status association with AQoL/SF-6D. ** $P=<0.05$. *** $P=<0.001$.

5D, it has not been validated in a critical illness population.

Data analysis

We used descriptive statistics to present the study findings. Scales for the AQoL and SF-6D were scored using the standard algorithms of the instrument developers. We report means and standard deviations or 95% confidence intervals for data which were normally distributed; otherwise we report the median with the interquartile range (IQR). We used the intraclass correlation co-efficient (ICC) (two-way mixed-model set for absolute agreement) to analyse agreement between the AQoL and SF-6D.

We assessed reliability with Cronbach's α, which should not fall below 0.70 and should exceed 0.80,[23] and we assessed sensitivity through comparisons between known groups, using independent *t* tests or analysis of variance (ANOVA). We assessed responsiveness by examining the change in MAU over time (before ICU to after ICU) using the dependent *t* test. Effect sizes (ESs) over time were defined according to Kazis et al[24] and interpreted after Cohen (small, $\leqslant 0.4$; moderate, 0.5; large, 0.8).[25]

To compare measurement sensitivity or responsiveness, we used the relative efficiency (RE) statistic.[26] We constructed predictive models using logistic regression and odds ratios (ORs) reported with 95% CIs. QALYs were calculated using the area-under-the-curve method.[27] We used SPSS version 15.0 (SPSS Inc) for statistical analysis, and $P<0.05$ was accepted for statistical significance.

Results

One hundred patients (82% of eligible participants) were initially recruited (Figure 1). Fifteen patients (15%) had died 6 months after ICU

Table 3. Responsiveness of the AQoL and SF-6D to changes in self-reported health status

	Pre-ICU and 6-month follow-up utility scores			
	AQoL*		SF-6D	
Panel 1. Change scores, baseline to follow-up after ICU stay†	**Mean**	**SD**	**Mean**	**SD**
Baseline data collection ($n = 67$)	0.49	0.32	0.57	0.14
Follow-up data collection ($n = 67$)	0.53	0.33	0.63	0.14
Utility change	0.05	0.29	0.06	0.15
t test (paired)	$t = -1.30$	–	$t = -3.09$**	–
Relative efficiency	1.00	–	5.65	–
Panel 2. Estimated responsiveness by change in health status, baseline to follow-up after ICU stay‡	**Mean change score**	**95% CI**	**Mean change score**	**95% CI**
No change ($n = 22$)	0.04	0.19	0.02	0.11
1-point absolute change in health status ($n = 34$)	0.20	0.18	0.11	0.09
t test (independent)	$t = -3.27$**	–	$t = -2.96$**	–
Relative efficiency	1.00	–	0.82	–

AQoL = Assessment of Quality of Life. SF-6D = Short Form 6D. ICU = intensive care unit.
* AQoL was set as the denominator for all tests, for ease of understanding. † Pooled analyses.
‡ Absolute analyses (regardless of improvement/deterioration state). ** $P < 0.01$.

discharge, leaving 85 participants eligible for follow-up. Six participants (6%) were lost to follow-up and two (2%) refused follow-up. Ten remaining participants had proxy-completed questionnaires and were excluded from data analysis, leaving 67 patients in the study (Table 1). At 6-month follow-up, 64/67 of these (96%) had returned to their preadmission place of residence, and 5/17 patients (29%) who had been in paid employment before their admission to the ICU had returned to work.

Utility distribution and reliability

The mean health utility scores of the sample at baseline were 0.49 (SD, 0.32) as measured by the AQoL, and 0.57 (SD, 0.14) as measured by the SF-6D. At 6-month follow-up they were 0.53 (SD, 0.33) and 0.63 (SD, 0.14), respectively. The median QALYs at 6 months, as measured by the AQoL and SF-6D respectively, were 0.27 (IQR, 0.14–0.38) and 0.31 (IQR, 0.27–0.35). There was moderate agreement at baseline between AQoL and SF-6D utilities ($ICC_{2,1} = 0.71$; 95% CI, 0.50–0.83). There was no significant ceiling or floor effect for either instrument. Figure 2 shows the AQoL and SF-6D frequency distributions at baseline. However, at both baseline and follow-up, one item of the SF-6D (role function) was observed to have no variance. The Cronbach αs for the AQoL at baseline and follow-up were 0.81 and 0.82 respectively, and for the SF-6D at baseline and follow-up were 0.65 and 0.70 respectively.

Sensitivity

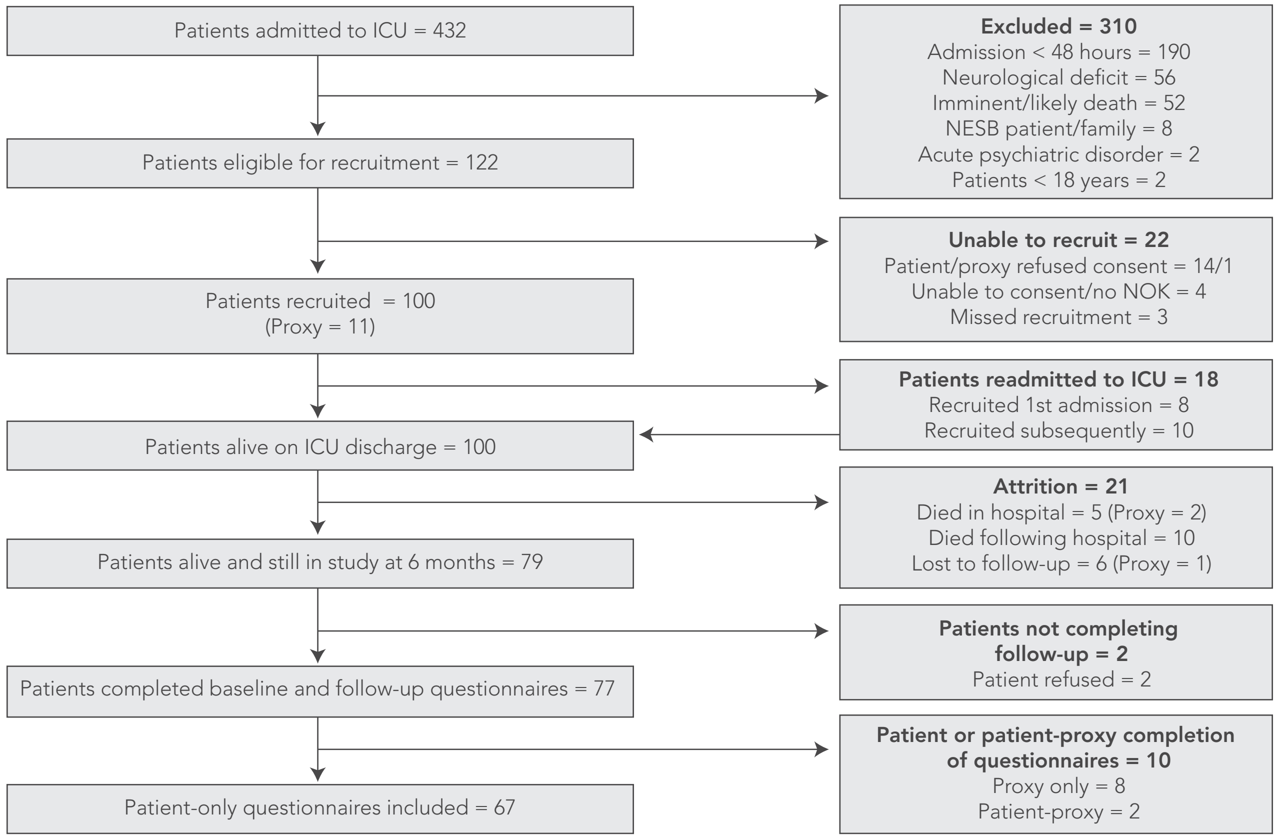
Both instruments were sensitive at both time points for general health status and rehabilitation status at follow-up (Table 2). The AQoL was sensitive to an increased requirement for community health services, but the SF-6D was not. Neither instrument varied by sex, age, or discharge destination. Of the four available RE tests, the SF-6D was more efficient than the AQoL for 6-month rehabilitation status. The AQoL was more efficient for home discharge status and for 6-month follow-up general health status. There was no difference by health status at baseline (Table 2).

Responsiveness

Forty-eight per cent of participants reported deterioration in utility using the AQoL (measured by change scores between baseline and follow-up), compared with 34% using the SF-6D. The difference between pre-ICU and post-ICU scores was modest, at 0.05 utilities for the AQoL and 0.06 for the SF-6D (Table 3). When expressed as Kazis's ES,[24] small ESs were observed (AQoL ES, 0.13 [95% CI, 0.05–0.20]; SF-6D ES, 0.43 [95% CI, 0.40–0.46]). The difference in the statistical tests was due to the differences in the SDs, with the SF-6D having an SD about half that of the AQoL. The RE statistic suggests that the SF-6D was five times more responsive than the AQoL in this sample.

Finally, the baseline AQoL and SF-6D utilities were used to predict postdischarge readmission to hospital or the ICU, or attendance at the emergency department (categorised as "readmitted"). AQoL and SF-6D utility scores were dichotomised at 1 SD below the population norm; for the AQoL this was at 0.58 and for the SF-6D was at 0.64. Logistic regression was used to examine whether participants with AQoL or SF-6D utility scores below these cut-off points were more likely to have been readmitted. Two potential confounders were included in the model: sex of the participant and whether the reasons for ICU admission were primarily medical or surgical (in the models, neither of these potential confounders was statistically significant). The results showed that participants obtaining AQoL scores more than 1 SD

Figure 1. Patient movement through the study period



ICU = intensive care unit. NESB = non-English speaking background. NOK = next of kin.

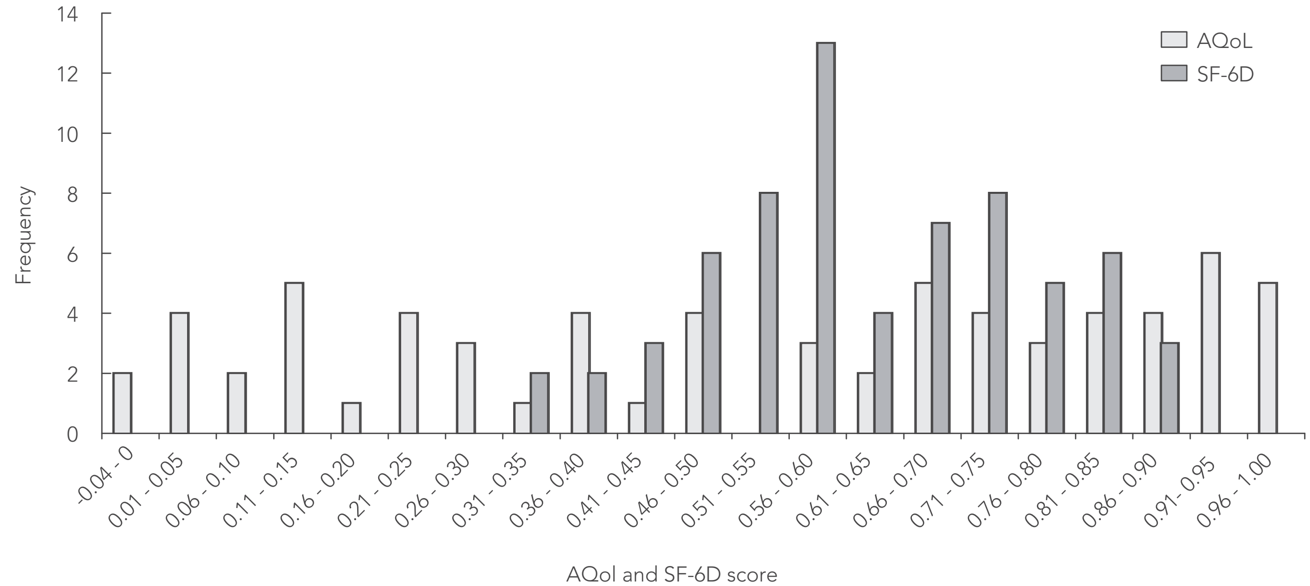
below the population norm had three times the OR of being readmitted, compared with those with AQoL scores less than 1 SD below the population norm (OR, 3.05; 95% CI, 1.00–9.28; $P = 0.049$) (correct classification, 70.1%; Hosmer–Lemeshow goodness-of-fit test $\chi^2 = 10.16$; $P = 0.04$; Nagelkerke's $R^2 = 13.3\%$). For the SF-6D, there was no statistically significant prediction between scores and readmission (OR, 1.61; 95% CI, 0.55–4.74; $P = 0.41$) (correct classification, 62.7%; Hosmer–Lemeshow goodness-of-fit test $\chi^2 = 3.69$, $P = 0.72$; Nagelkerke's $R^2 = 7.0\%$).

Discussion

The demographic and clinical features of the sample were broadly representative of a typical mixed medical and surgical ICU population,[28,29] although the median Acute Physiology and Chronic Health Evaluation II score reflected the fact that the sample was of a relatively low acuity. Previous research suggests that women and older adults report worse HRQoL scores,[19,30,31] but no differences were found in utility scores by sex or age, and it is possible that the impact of critical illness (or the contributing factors to critical illness) dominates sex and age differences in the ICU population. Therefore, caution should be exercised when testing instrument validity against demographic factors in the critical illness population.[32,33]

The key difference between the AQoL and SF-6D utilities was the difference in distribution. The AQoL demonstrated an even distribution across the utility scale of 0.00 to 1.00, compared with the SF-6D, which showed a bell-shaped curve around 0.60. This clustering of SF-6D utility scores has been noted in other study populations, with the lower boundary score (0.30) a factor.[34] These findings have implications for the use of the SF-6D in this population and may lead to an underestimation of the QALYs gained for use in CUAs. This is particularly so for populations in which patients have poor outcomes, worse pre-existing disease or poorer preadmission HRQoL scores. QALYs are an outcome measurement unit which combines quality of life and duration of survival. In CUA, "QALY gain" refers to the increase in QALYs per dollar cost of treatment, and

Figure 2. AQoL and SF-6D frequency distribution (*n* = 67)



AQoL = assessment of quality of life. SF-6D = Medical Outcomes Study Short Form 6D.

treatment and policy decisions are based on the QALY gain per dollar.

In our study, both the AQoL and SF-6D instruments showed mean scores that were well under population norms. For the AQoL, the Australian population norm is 0.81 (SD, 0.22),[35] suggesting that participants experienced a HRQoL that was 0.32 utility points worse than the general population. Similarly for the SF-6D, the Australian population norm (G H, own unpublished data) is 0.78 (SD, 0.14), suggesting that participants experienced a HRQoL that was 0.21 utility points worse than the general population. When used in CUA, these figures imply that for a 1-QALY gain, studies using the AQoL would require only three patients to return to the population norm, compared with studies using the SF-6D which would require five patients to return to the population norm for the equivalent QALY gain. The QALYs of this cohort 6 months after admission to the ICU were higher than survivors of acute respiratory distress syndrome,[36] which is to be expected, given the higher severity of illness of that cohort.

Another key difference between the AQoL and SF-6D utilities was that the AQoL SDs were more than double the SF-6D SDs, and that participants in our study reported an SD of 0.32 in AQoL utility (compared with the population norm SD of 0.22[35]). These differences may be a function of the richer descriptive system of the AQoL. Additionally, the finding of an SD of 0.32 in AQoL utility may reflect the fact that patients have a broad experience of, and react differently to, life-threatening illness. In contrast, the SF-6D SD was the same as the population SD (0.14). It is possible that this partly reflects the lack of variance in the role function composite item of the SF-6D, as well as the more compressed utility range of the SF-6D. These observations suggest that the SF-6D may provide incorrect estimates of benefit from health care, and hence may also provide misleading cost–utility ratios, as has been reported previously.[34] Despite the moderate agreement between the two utility scores, the absolute utility differences (by health status at baseline and follow-up) were large, with differences on the AQoL being about 50% greater than those reported on the SF-6D. These differences are clinically significant. This finding suggests that when comparing different measures, correlational analysis is an inadequate procedure by itself, as has been previously suggested by Bland and Altman.[37] These key differences and the differences in psychometric properties observed support the conclusion that utilities from different instruments are not directly comparable, equivalent or able to be used interchangeably.[11,34,38] This has significant implications for MAU instrument selection for CUA in clinical studies in the critical illness population.

At preadmission and follow-up, the AQoL was a reliable measure, and this is consistent with reports in other populations.[3,20] However, important reliability issues were observed for the SF-6D where one item had no variance, suggesting that the SF-6D did not meet the standards for reliability and should be used with caution in this population. The sensitivity of the AQoL and SF-6D varied according to the scenario being considered. The AQoL outperformed the SF-6D in terms of sensitivity to activities that are clinically important to patients and health care systems, such as the requirement for community services. The AQoL also outperformed the SF-6D for people with poor general health and for people who required support at home. Conversely, where the SF-6D outperformed the AQoL, the differences were largely due to the smaller SD of the SF-6D. These findings may also reflect differences in instrument structure, with the AQoL measuring handicap arising from an intrinsic health state,[3] while the SF-6D was designed to assess actual health states.[6]

Although both instruments were responsive, only small changes in mean utility scores were observed (AQoL change, 0.05; SF-6D change, 0.06) (Table 3). These small aggregate changes might suggest that participants may have to experience significant improvement (or decrement) in health utility before changes are considered clinically worthwhile or significant. There are several possible reasons for this. There is no gold standard with which to compare change after critical illness. It is also possible that the small utility changes were a function of the wide variability in outcome, where almost half the participants deteriorated during the study period and subgroups of patients had variable outcomes.[17]

The test of predictive validity suggested that those obtaining AQoL scores lower than 1 SD below the population norm were over three times more likely to re-present at hospital, be readmitted to the ICU, or present at the emergency department. The clear clinical implication from this finding is that patients with AQoL scores lower than 1 SD below the population norm should be carefully screened or monitored over time for deterioration in their symptoms. The fact that there was no statistically significant prediction for readmission using the SF-6D adds to the weight of evidence suggesting that it may not be particularly suitable for use in studies of people in intensive care.

Our study was limited by the small sample size and an inability to generalise results, as only 23% of all admitted ICU patients were eligible and recruited for it. Interpreting the results of our study requires an acceptance that no consensus exists about the best method of deriving utility scores (eg, scaling preference-based instruments, or mapping or calculating scoring algorithms from health profile data).[39] However, instrument validity is vitally important, as the differences in utility scores may influence clinical decision making, resource allocation and CUA. Only two MAU instruments were used in the current study, for pragmatic reasons; future studies could include other MAU instruments such as the HUI and EQ-5D. Future validation of MAU instruments could also consider validating perception of patient change against family or caregiver burden, which has been described in the critical illness literature[40] and may represent an important part of the paradigm of ICU outcome.

Conclusions

The AQoL showed better psychometric properties than the SF-6D in this sample of Australian ICU survivors, and also showed acceptable internal reliability. The SF-6D may not meet contemporary standards of reliability, and the lower boundary score of 0.30 may limit the use of the SF-6D in this population. The current advantage of the SF-6D is that it is derived from the SF-36, which has a wide utilisation. Moderate agreement was observed between the utility scores of the AQoL and SF-6D, but their utilities were not interchangeable. The AQoL is recommended for use in the critical illness population.

Competing interests

None declared.

Author details

Elizabeth H Skinner, Senior Physiotherapist, Intensive Care,[1] and Honorary Research Fellow[2]
Linda Denehy, Head of Physiotherapy[2]
Stephen Warrillow, Deputy Director, Intensive Care[3,4]
Graeme Hawthorne, Principal Research Fellow[5,6]

1 Departments of Physiotherapy and Intensive Care, Southern Health, Melbourne, VIC, Australia.
2 Department of Physiotherapy, School of Health Sciences, University of Melbourne, Melbourne, VIC, Australia.
3 Melbourne School of Health Sciences, University of Melbourne, Melbourne, VIC, Australia.
4 Department of Surgery, Austin Health, University of Melbourne, Melbourne, VIC, Australia.
5 Mental Health Evaluation Unit, Department of Psychiatry, University of Melbourne, Melbourne, VIC, Australia.
6 Northern Health Research Centre, Department of Medicine, University of Melbourne, Melbourne, VIC, Australia.

Correspondence: elizabeth.skinner@wh.org.au

References

1 Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.
2 Santana MJ, Feeny DH. Using the health utilities index in routine clinical care: process, feasibility, and acceptability: a randomized controlled trial. *Patient* 2009; 2: 159-67.
3 Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health-related quality of life. *Qual Life Res* 1999; 8: 209-24.
4 Feeny D, Furlong W, Barr RD, et al. A comprehensive multiattribute system for classifying the health status of survivors of childhood cancer. *J Clin Oncol* 1992; 10: 923-8.

5 EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16: 199-208.

6 Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol* 1998; 51: 1115-28.

7 Orwelius L, Nordlund A, Nordlund P, et al. Pre-existing disease: the most important factor for health related quality of life long-term after critical illness: a prospective, longitudinal, multicentre trial. *Crit Care* 2010; 14: R67.

8 Iwashyna TJ, Ely EW, Smith DM, Langa KM. Long-term cognitive impairment and functional disability among survivors of severe sepsis. *JAMA* 2010; 304: 1787-94.

9 Talmor D, Shapiro N, Greenberg D, et al. When is critical care medicine cost-effective? A systematic review of the cost-effectiveness literature. *Crit Care Med* 2006; 34: 2738-47.

10 Halpern NA, Pastores SM. Critical care medicine in the United States 2000-2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. *Crit Care Med* 2010; 38: 65-71.

11 Marra CA, Marion SA, Guh DP, et al. Not all "quality-adjusted life years" are equal. *J Clin Epidemiol* 2007; 60: 616-24.

12 Schackman BR, Gold HT, Stone PW, Neumann PJ. How often do sensitivity analyses for economic parameters change cost-utility analysis conclusions? *Pharmacoeconomics* 2004; 22: 293-300.

13 Hawthorne G, Richardson J. Measuring the value of program outcomes: a review of multiattribute utility measures. *Expert Rev Pharmacoecon Outcomes Res* 2001; 1: 215-28.

14 Kaarlola A, Tallgren M, Pettilä V. Long-term survival, quality of life, and quality-adjusted life-years among critically ill elderly patients. *Crit Care Med* 2006; 34: 2120-6.

15 Rosenberg AL, Tremper KK. Wide variations in the cost-effectiveness of critical care. *Crit Care Med* 2006; 34: 2849-51.

16 Vainiola T, Pettilä V, Roine RP, et al. Comparison of two utility instruments, the EQ-5D and the 15D, in the critical care setting. *Intensive Care Med* 2010; 36: 2090-3.

17 Skinner EH, Warrillow S, Denehy L. Health-related quality of life in Australian survivors of critical illness. *Crit Care Med* 2011; 39: 1896-905.

18 Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Boston: The Health Institute, 1993.

19 Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002; 21: 271-92.

20 Osborne RH, Hawthorne G, Lew EA, Gray LC. Quality of life assessment in the community-dwelling elderly: validation of the Assessment of Quality of Life (AQoL) Instrument and comparison with the SF-36. *J Clin Epidemiol* 2003; 56: 138-47.

21 Hawthorne G, Richardson J, Day N. Using the Assessment of Quality of Life (AQoL) Instrument. Version 1.0. Technical report 12. Melbourne: Centre for Health Economics, 2000.

22 Black NA, Jenkinson C, Hayes JA, et al. Review of outcome measures used in adult critical care. *Crit Care Med* 2001; 29: 2119-24.

23 DeVellis RF. Scale development: theory and applications. 2nd ed. Thousand Oaks, CA: Sage, 2003.

24 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989; 27 (3 Suppl): S178-89.

25 Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.

26 Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 1985; 28: 542-7.

27 Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. *BMJ* 1990; 300: 230-5.

28 Cuthbertson BH, Scott J, Strachan M, et al. Quality of life before and after intensive care. *Anaesthesia* 2005; 60: 332-9.

29 Elliott D, Mudaliar Y, Kim C. Examining discharge outcomes and health status of critically ill patients: some practical considerations. *Intensive Crit Care Nurs* 2004; 20: 366-77.

30 Hawthorne G, Osborne R. Population norms and meaningful differences for the Assessment of Quality of Life (AQoL) measure. *Aust N Z J Public Health* 2005; 29: 136-42.

31 Lipsett PA, Swoboda SM, Campbell KA, et al. Sickness Impact Profile Score versus a Modified Short-Form survey for functional outcome assessment: acceptability, reliability, and validity in critically ill patients with prolonged intensive care unit stays. *J Trauma* 2000; 49: 737-43.

32 Kirchengast S, Haslinger B. Gender differences in health-related quality of life among healthy aged and old-aged Austrians: cross-sectional analysis. *Gend Med* 2008; 5: 270-8.

33 Orpana HM, Ross N, Feeny D, et al. The natural history of health-related quality of life: a 10-year cohort study. *Health Rep* 2009; 20: 29-35.

34 Hawthorne G, Richardson J, Day NA. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Ann Med* 2001; 33: 358-70.

35 Hawthorne G, Korn S, Richardson J. Population norms for the AQoL derived from the 2007 Australian National Survey of Mental Health and Wellbeing. *Aust N Z J Public Health* 2013; 37: 7-16.

36 Ferguson ND, Scales DC, Pinto R, et al. Integrating mortality and morbidity outcomes: using quality-adjusted life years in critical care trials. *Am J Respir Crit Care Med* 2013; 187: 256-61.

37 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-10.

38 Holland R, Smith RD, Harvey I, et al. Assessing quality of life in the elderly: a direct comparison of the EQ-5D and AQoL. *Health Econ* 2004; 13: 793-805.

39 Mortimer D, Segal L. Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights. *Med Decis Making* 2008; 28: 66-89.

40 Van Pelt DC, Milbrandt EB, Qin L, et al. Informal caregiver burden among survivors of prolonged mechanical ventilation. *Am J Respir Crit Care Med* 2007; 175: 167-73. ❑