UTS:CHERE

CHERE

CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

UNIVERSITY OF TECHNOLOGY SYDNEY

The Centre for Health Economics Research and Evaluation (CHERE) was established in 1991. CHERE is a centre of excellence in health economics and health services research. It is a joint Centre of the Faculties of Business and Nursing, Midwifery and Health at the University of Technology, Sydney, in collaboration with Central Sydney Area Health Service. It was established as a UTS Centre in February, 2002. The Centre aims to contribute to the development and application of health economics and health services research through research, teaching and policy support. CHERE's research program encompasses both the theory and application of health economics. The main theoretical research theme pursues valuing benefits, including understanding what individuals value from health and health care, how such values should be measured, and exploring the social values attached to these benefits. The applied research focuses on economic and the appraisal of new programs or new ways of delivering and/or funding services. CHERE's teaching includes introducing clinicians, health services managers, public health professionals and others to health economic principles. Training programs aim to develop practical skills in health economics and health services research. Policy support is provided at all levels of the health care system by undertaking commissioned projects, through the provision of formal and informal advice as well as participation in working parties and committees.

University of Technology, Sydney
City campus, Haymarket
PO Box 123 Broadway NSW 2007
Tel: +61 2 9514 4720
Fax: + 61 2 9514 4730
Email: mail@chere.uts.edu.au
www.chere.uts.edu.au

Validity of the SF-36 Health Survey as an outcome measure for trials in people with spinal cord injury

Mark J Haran[1], Madeleine T King[2], Martin R Stockler[3] Obad Marial [4] and Bonne B Lee[4]

Corresponding Author: Dr Mark J Haran
Department of Aged Care and Rehabilitation Medicine
Building 12, Royal North Shore Hospital
Pacific Highway, St Leonards NSW Australia 2065
Ph. 612-9926-8705 Fax. 612-9906-4301
Email mharan@med.usyd.edu.au

CHERE WORKING PAPER 2007/4

1. Royal North Shore Hospital, Sydney

2. Centre for Health Economics Research and Evaluation
Faculty of Business
University of Technology, Sydney

3. NHMRC Clinical Trials Centre
University of Sydney
Sydney Cancer Centre – Royal Prince Alfred and Concord Hospitals

4. Prince of Wales Hospital, Sydney

First Version: June 2004
Current Version: June 2007

ABSTRACT

STUDY DESIGN: Cross-sectional analytical study.

OBJECTIVES: To assess the measurement properties of the Medical Outcomes Study Short-Form 36 (SF-36) Health Survey (Standard version) for people with spinal cord injury.

SETTING: Predominantly community-based sample living in New South Wales, Australia, recruited to a randomised trial assessing the efficacy of urinary antiseptics.

METHODS: The SF-36 was interviewer-administered to 305 subjects at recruitment. Feasibility, content validity and internal consistency were assessed. We tested a priori hypotheses about discriminative, convergent and divergent validity.

RESULTS: Interviewer-assisted administration was feasible. The content validity of several domains (Physical Function, Role Physical, Social Function and Role Emotional) was compromised by the irrelevance of some items and response options. Resultant ceiling and floor effects may limit the SF-36's ability to detect changes over time. The SF-36 was able to discriminate differences between people with: tetraplegia versus paraplegia (in the Physical Function and Physical Composite scores); injuries that were recent (<4 years) versus remote (>4 years) (in the Vitality, Social Function and Mental Health domain and Mental Composite scores), and who were employed versus unemployed (in the Physical Function, Social Function, Mental Health and Mental Composite scores). It was not able to discriminate between groups dichotomised by age, injury completeness or gender. The convergent and divergent validity of all SF-36 domains was as in other populations, except for correlations involving the Physical Function scale which were poor. Internal consistency was similar to that in other populations (Cronbach's alpha from 0.75 to 0.92); the SF-36 has sufficient precision for population-based and clinical research in spinal cord injury.

CONCLUSION: The SF-36 is useful for comparing the health status of people with spinal cord injury to that of other populations, but supplementation with a disease-specific health status measure may be necessary for trials of interventions in people with spinal cord injuries.

ACKNOWLEDGEMENTS

This study was supported by a grant provided by the Motor Accidents Authority of New South Wales.
.

KEYWORDS

Health Status, Quality of life, Spinal Cord Injury, validity, SF-36

ABBREVIATIONS

SF-36	Short-Form 36 Health Survey
PF	Physical Functioning domain of SF-36 Health Survey
RP	Role limitation due to Physical problems (Role-Physical) domain of SF-36 Health Survey
BP	Bodily Pain domain of SF-36 Health Survey
GH	General Perception of Health (GH) domain of SF-36 Health Survey
VT	Vitality domain of SF-36 Health Survey
SF	Social Function (SF) domain of SF-36 Health Survey
RE	Role limitation due to Emotional problems (Role-Emotional) domain of SF-36 Health Survey
MH	Mental Health domain of SF-36 Health Survey
PCS	Physical Composite Scale of SF-36 Health Survey
MCS	Mental Composite Scale of SF-36 Health Survey
RNSH	Royal North Shore Hospital
POWH	Prince of Wales Hospital
RCT	Randomised Controlled Trial
ASIA	American Spinal Injury Association
ES	Effect size.

INTRODUCTION

Health status and health-related quality of life are frequently used synonymously. They represent abstract, multi-dimensional constructs including physical, psychological, and social aspects, which include, but are not limited to, the concept of health [1-3], and which reflect an individual's perception of, and response to, their unique circumstances [3-4]. They are dynamic concepts that may mean different things to different people at different times in their lives [5].

Spinal cord injury profoundly impacts health status [6-8]. As a group, people with spinal cord injury suffer high rates of suicide, self-neglect, divorce and drug abuse [9-11], and face prolonged psychosocial adjustment, enforced lifestyles changes, and frequent medical complications [12]. Yet many report good health status [9-11], and patients' ratings of their own physical and mental functioning differ substantially from those of their treating clinicians [9,13]. Clinical trials of interventions in this population should therefore incorporate patient-reported measures of health status.

The SF-36 is a generic, brief, multi-dimensional, self-report health questionnaire that measures eight health concepts - Physical Functioning (PF), Role limitation due to Physical problems (Role-Physical, RP), Bodily Pain (BP), General Perception of Health (GH), Vitality (VT), Social Function (SF), Role limitation due to Emotional problems (Role-Emotional, RE), and Mental Health (MH). The domain scales ranges from 0 (worst possible health state measured by the questionnaire) to 100 (best possible health state). The domain scales can be aggregated into two composite scales, physical (PCS) and mental (MCS) [14]. which are standardised to a mean score of 50 and standard deviation of 10. The SF-36 scales help describe differences between populations in physical and mental health status, burden of chronic disease, and multi-dimensional effects of interventions on health status [14-20], but their usefulness for screening for health problems in individual patients is questionable [21].

The SF-36 has demonstrated validity and internal consistency when used in the general populations [22-29]. The SF-36's construct validity is supported by its ability to discriminate between people with and without various physical and mental health problems [14,22,23]. Recently, SF-36 data has been published for 587 Canadians who suffered a spinal injury two or more years earlier, showing that all domain and physical composite scores (but not mental composite scores) were significantly lower than those of the American general population [12]. Similar age- and gender-adjusted SF-36 scores for Australians with spinal cord injury and neurogenic bladder have also recently been published [30]. These demonstrate the utility of the SF-36 for burden of illness studies. But to be useful in clinical trials, the SF-36 must be able to discriminate at a finer level of detail among subgroups of people with spinal cord injury. This aspect of the validity of the SF-36 has not been assessed in Australians with spinal cord injury.

The validity of any health status measure should be examined in each new context [3,17]. The aim of this study was to assess the validity of the SF-36 as an outcome measure for clinical trials in Australians with spinal cord injury. It is an example of validation by application using baseline data from a randomised trial.

METHODS

Subjects were sampled from a comprehensive, composite register, comprising the New South Wales Spinal Cord Injuries Database [31] and the admissions records for the only two acute spinal services in New South Wales (Royal North Shore Hospital (RNSH) [32] and Prince of Wales Hospital (POWH)).

Subjects were predominantly community-based consenting participants in a randomised controlled trial (RCT) of antiseptic agents for the prevention of urinary infections in persons with spinal cord injury and neurogenic bladder (Lee BB et al; article accepted for publication by Archives of Physical Medicine and Rehabilitation in August 2006). The trial was approved by the ethics committees of the participating hospitals (RNSH, POWH, Prince Henry Hospital, and Royal Rehabilitation Centre Sydney). Inclusion criteria for the trial were: spinal cord injury with neurogenic bladder; stable bladder management with either intermittent catheterisation, indwelling urethral or suprapubic catheter, or reflex voiding with or without condom drainage; absence of complex urological or serious renal pathology; not being prescribed antibiotics at the time of enrolment, and; an absence of current symptoms of a urinary tract infection.

Between November 2000 and August 2002 a sequential sample of 543 eligible people were contacted, and 305 (56%) agreed to participate. This provided the requisite sample for the RCT, and represented 9% of the composite register. Subjects completed the Standard version of the SF-36 health questionnaire [14] when they were recruited to the study. Most completed the SF-36 themselves with a research officer present (55% of subjects). Subjects who were more physically-impaired responded to a face-to-face interview (42%), while 3% completed the SF-36 at home and returned it by mail. Incomplete responses were rectified by direct enquiry or telephone follow-up as necessary. Interpreters were used where necessary (1%).

Individual SF36 items were coded, summed and transformed for each patient into the eight domain scales [14,33]. The physical and mental composite scores were calculated using Australian factor weightings [22].

We assessed the following measurement properties: feasibility of administration; content validity; discriminative validity; convergent and divergent validity; and internal consistency.

FEASIBILITY

Feinstein [34] considered questionnaire design and ease of use as an aspect of validity. Practical difficulties with the SF-36 in this population were assessed from the recorded comments of the subjects, research assistant and authors (OM, BL, MH) during administration. Completion rates were recorded.

CONTENT VALIDITY

Content validity refers to how comprehensively the SF-36 items cover the key concepts, or how adequately they reflect the aims of the index, in this case to measure health status in a population with spinal cord injury [35]. Content validity was assessed from the experience of the clinician researchers (MH, BL, OM) during the administration of the SF-36.

CONSTRUCT VALIDITY

Construct validity is the extent to which the questionnaire supports predefined hypotheses about expected relationships among domains (within and between instruments) and with other clinically-relevant measures [35]. In this study three types of construct validity were assessed: discriminative, convergent and divergent validity.

Discriminative Validity

Discriminative validity refers to the ability of a measure to discriminate among groups of individuals whose health status is expected to differ [35]. We have previously demonstrated [30] the SF-36's ability to discriminate across most domains between those with and without spinal cord injury [22], but this is a very coarse distinction.

At a more refined level, discriminative validity would be evident if the domain and composite scores of the SF-36 were able to discriminate among subgroups of people with spinal cord injury. We tested for differences between groups dichotomized by three clinical and three demographic variables: extent of impairment (tetraplegic versus paraplegic) [36]; completeness of injury (complete [ASIA classification A] versus incomplete [ASIA classification B, C or D]) [36]; time since spinal cord injury (<4 versus 4+ years); age <44 versus 44+ years old; gender; and employment status (employed versus unemployed) (see Table 1). We generated *a priori* hypotheses about the relationship between health status and these dichotomized variables; three investigators with expertise in either spinal cord injury (BL) or health status assessment (MK, MS) independently rated whether they expected the health status of groups dichotomized by these variables to differ to a substantial and consistent degree. They rated the expected differences as large, medium or small. Twenty-three of the 60 possible differences were expected to be medium or large by all 3 investigators; these were tested as *a priori* hypotheses, and are displayed in the shaded cells of Table 2. (The remaining 37 differences in Table 2 were not expected to be substantial and consistent.)

Table 1. Distribution of total sample (n=305) across dichotomised clinical criteria for the assessment of discriminative validity

Criterion Variables	**Worse Category**		**Better Category**	
	Criterion	***n (%)***	**Criterion**	***n (%)***
Tetraplegia/ Paraplegia	Tetraplegic	167 (55)	Paraplegic	138 (45)
Completeness of injury	Complete	148 (49)	Incomplete	157 (51)
Time since injury	≤ 4 years	90 (30)	>4 years	215 (70)
Age	16-43 years	157 (51)	44+ years	148 (49)
Gender	Female	53 (17)	Male	252 (83)
Employment	No paid hours/week	197 (65)	Any paid hours/week	108 (35)

The discriminative validity of the SF-36 was assessed by testing the differences in unadjusted domain and composite physical and mental scores across groups as displayed in Table 2. The distributions of the domain scores were not normal [30], so differences were tested using the Mann-Whitney U test [37]. To account for multiple comparisons, the significance levels for the 23 hypothesis tests were adjusted after Hochberg [38] using a two-tailed global significance level of 0.05. For these comparisons, a significance level which reached the Hochberg cut-off was deemed to represent strong evidence of discriminative validity, and a significance level between the Hochberg value and 0.05 was deemed weak supportive evidence.

The clinical significance of these differences was interpreted in terms of effect sizes, calculated by dividing the difference between means by the sample standard deviation. We defined the clinical significance of effect sizes (ES) *a priori*, according to previously described criteria, as being small (0.2), moderate (0.5), or large (0.8) [39,40].

Convergent and Divergent Validity

Theory suggests that some domains of health status should be correlated (convergent validity), while others are anticipated to be relatively unrelated (divergent validity) [35]. Correlation coefficients between scales provide measures of convergent and divergent validity. Given that the domain scores were not normally distributed [30], Spearman's rank correlation coefficients were used [41-43]. The clinical significance of correlations was defined *a priori* as: <0.30 (weak), 0.3-0.5 moderate and >0.5 (strong) [41-43]. Based on patterns observed in other studies [17], we expected convergent validity to be evidenced by moderate to strong correlation among the three physical scales (PF, RP and BP), between the two general health scales (GH and VT) and among the three psychological scales (SF, RE and MH). All other correlations were expected to be weak to moderate, reflecting divergent validity.

INTERNAL CONSISTENCY

Internal consistency represents the extent to which the component items within a scale are correlated to one another [44]. This was evaluated by Cronbach's alpha, an inter-item correlation statistic with a range of 0 to 1 [44]. Higher values for Cronbach's alpha indicate that items on a dimension are related, such that the scale measures a single underlying variable with more precision [45]. Alpha values above 0.5 may be acceptable [46] but values above 0.8 [47] are recommended for instruments used to assess groups, and values above 0.9 are recommended for instruments to assess individuals [48].

Tests of hypotheses were performed with Minitab for Windows Release 12 [49], and Spearman's correlations and Cronbach's alpha calculations were calculated with SPSS Version 11.0 [50].

RESULTS

We approached 543 people to enter the study and 56% (n=305) consented to participate. The sample had a mean age of 44 (standard deviation [SD] 14) years (range 16-82 years) and most were male (83%). Fifty-five percent of patients had tetraplegia and 49% had a complete spinal injury. The mean time since spinal cord injury was 14 (SD12) years (range 1 month to 61 years).

FEASIBILITY

Some subjects with upper limb impairment had problems with writing and positioning the document. In such cases, the research assistant helped the subject. Using such methods, there was no missing data. It took subjects an average of 15 minutes to complete the SF-36.

CONTENT VALIDITY

Questions that were frequently directed to the research assistant during administration included: clarification of whether the questions relating to limitations of activities referred to a baseline comparison with the performance of the normal population or the patient's usual activities; difficulty answering questions relating to strenuous activities because patients indicated that they did not usually do such activities; and, uncertainty regarding the period of recall.

CONSTRUCT VALIDITY

Discriminative Validity

Evidence for the discriminative validity of the SF-36 in this population is presented in Table 2. There was strong evidence (p<Hochberg critical value) that the Physical Function and Physical Composite scales discriminated between people with paraplegia and those with tetraplegia; these differences were large (ES=1.1) and small to moderate (ES=0.39), respectively. Similarly, there was strong evidence that the Mental Health and Mental Composite scales discriminated between the employed and the unemployed; these differences were small to moderate (ES=0.37 and 0.30, respectively).

There was weak evidence (Hochberg critical value<p<0.05) that the Physical Function, Social Function and Vitality scales discriminated between the employed and the unemployed; these differences were small (ES=0.26, 0.30 and 0.23 respectively). There was equivocal evidence (p=0.06) that the General Health scale discriminated between the employed and the unemployed (ES=0.24) and that the Physical Composite scale discriminated between the young and the old (ES=0.22).

No other *a priori* expectations were confirmed by the data. However, of the remaining 37 comparisons, Bodily Pain scores were significantly worse in persons with paraplegia compared with tetraplegia, and the group that was more than 4 years post-injury had better Vitality, Social Function, Mental Health, and Mental Composite scores than those with more recent injuries.

Table 2. Discriminative validity^ of the SF-36 in the spinal cord injured population.

SF-36 scale	**Differences in mean (standard error) and *p* values* for SF-36 scores between groups dichotomized by:** Paraplegia Vs Tetraplegia	Incomplete vs Complete Injury	Injury 5+ vs 0-4 years ago	Age 16-43 vs 44+ years	Male Vs Female	Employed vs Unemployed
Physical function	**20.6 (2) <0.0001#**	4.4 (2)	0.2 (3)	-0.4 (2)	2.4 (2)	**5.0 (2) 0.008**
Role Physical	-2.7 (5)	-6.1 (5) 0.16	8.7 (5) 0.18	2.2 (5)	0.2 (6)	7.2 (5)
Bodily Pain	**-8.9 (3) 0.01**	-0.3 (3)	2.5 (4)	4.4 (3) 0.17	6.4 (5) 0.17	2.7 (3)
General Health	2.0 (3)	0.4 (3)	-0.1 (3)	-0.7 (3)	4.4 (3)	5.1 (3) 0.06
Vitality	2.7 (3)	-2.7 (2) 0.14	**6.3 (3) 0.01**	0.5 (3) 0.20	4.0 (3) 0.20	**4.8 (3) 0.05**
Social Function	-2.6 (3)	-2.1 (3)	**8.7 (4) 0.01**	-4.9 (3) 0.17	-3.5 (4) 0.19	**8.9 (3) 0.008**
Role Emotional	-3.7 (4)	1.7 (4)	5.2 (5)	-4.8 (4) 0.13	9.9 (6) 0.13	5.8 (4)
Mental Health	-0.2 (2)	-1.4 (2)	**5.9 (3) 0.005**	-4.2 (2)	4.2 (3)	**7.1 (2) 0.0004#**
Physical Composite	**3.0 (1) 0.0005#**	0.4 (1)	0.1 (1)	1.7 (1) 0.06	0.7 (1)	1.2 (1) 0.14
Mental Composite	-2.0 (1) 0.08	-1.0 (2) 0.12	**3.7 (2) 0.008**	-2.5 (1) 0.13	2.9 (2) 0.07	**3.6 (1) 0.003#**

^ Comparing the group expected to have the best health status with the group expected to have the worst health status

KEY: *A priori* expectations of strong and consistent differences across criterion are represented in the shaded cells. Within each cell the differences in mean scores (standard errors) are displayed on the top line and the *p* value on bottom line. *P* values >0.2 are not displayed.

* Mann Whitney U test.

P value reaches Hochberg significance level.

Convergent and divergent validity

The correlation matrix for the SF-36 domain scales is shown in Table 3. Convergent validity was generally supported; five of the seven predicted correlations were at least moderate, but the correlations between PF and both RF and BP were not. Some of the remaining correlations were stronger than expected, particularly those involving the Vitality, Social Function and Mental Health scales. The lowest correlations were between the Physical Function scale and all other scales.

Table 3. Spearman's Correlation Coefficients Testing Associations Between SF-36 Domain Scores.

	PF	RP	BP	GH	VT	SF	RE
RP	0.03						
BP	0.04	**0.31**					
GH	0.21	0.25	0.36				
VT	0.17	0.46	0.48	**0.48**			
SF	0.10	0.50	0.46	0.41	0.60		
RE	0.04	0.34	0.30	0.29	0.45	**0.43**	
MH	0.15	0.35	0.46	0.49	0.66	**0.58**	**0.51**

KEY: *A priori* expectations of moderate to high positive correlations (convergent validity) are represented in the shaded cells.

INTERNAL CONSISTENCY

The Cronbach's alpha scores ranged from 0.75 for the General Health domain to 0.92 for the Role-Emotional domain (Table 4).

Table 4. Cronbach's Alpha Statistics For The SF-36 Domains ((n=305)

SF-36 Domain Scale	Cronbach's Alpha Statistic
Physical Functioning	0.83
Role – Physical	0.90
Bodily Pain	0.88
General Health	0.75
Vitality	0.81
Social Function	0.82
Role – Emotional	0.92
Mental Health	0.84

DISCUSSION

Our study examined the measurement properties of the SF-36 in a group of predominantly community-based people with spinal cord injury and neurogenic bladder living in New South Wales, Australia. High completion rates were possible with a research assistant present during administration but many items in the PF, RP, SF and RE domains were either irrelevant or had unsuitable response options. We have recently demonstrated that the SF-36 discriminates between people with and without spinal cord injury [30]. However, in the current study we have shown that the SF-36 was may be limited in its ability to detect more subtle but important differences in health-status among people with spinal cord injury. The convergent validity of the SF-36 was as expected [17], except for correlations involving the Physical Function domain. Internal consistency was satisfactory. The PF scale compromised the SF-36's feasibility, and content, convergent and divergent validity. Overall, our results suggest that the SF-36 may be useful in population-based studies (e.g. for burden of illness studies), but may not be sufficient for evaluating the impact of interventions or for managing individual patients.

Strengths of our study include the completeness of data and the generation of *a priori* hypotheses to assess discriminative validity. However, our sample is less heterogenous and less healthy than the whole population of people with spinal cord injury, because all our subjects had a neurogenic bladder. Continence problems have been associated with poorer scores for General Health, but not for other domains of the SF-36, in people with spinal cord injury [51]. While not strictly representative of all patients with spinal cord injury, our sample is similar in many respects to persons with spinal injury from numerous countries [12,32,52,53], and so it is suitable for the purpose of validation. Our analysis contributes to the understanding of health status measurement in patients with spinal cord injury and the planning of future studies.

FEASIBILITY

Assistance in questionnaire administration by interviewers reduced missing data rates but was labour intensive. Computerised administration may be an alternative to interviewer-assistance but is unlikely to match our high completion rates.

CONTENT VALIDITY

The SF-36 has a number of items that may be problematic for the people most disabled by their spinal cord injury, such as items relating to vigorous activities, climbing stairs, and walking in the Physical Function domain. As noted by Tate et al [5], many current measures of health status have such limitations when used in people with spinal injury. The profound physical disabilities seen in spinal cord injury combined with the focus on more strenuous activities and the limited response options in the Physical Function domain are likely to explain the pronounced floor effects (29%) [30]. Inappropriate or ambiguous content can lead to inconsistency among people in the interpretation of questions and an inability of the scales to register changes in health status over time.

A number of issues that are likely to be relevant to health status in our sample [6,12,51,54,55] are not included in the SF-36, such as sexuality, mental functioning, social participation, vision, feeding, hospitalisations and illnesses, recreation and hobbies, continence, wheelchair mobility, and communication. Sexuality correlates poorly with SF-36 scales and is likely to add substantially to measures of health in many populations [17].

Another problem with the content validity of the SF-36 in this population is the ambiguity of what is meant by ‘usual activities’, as noted elsewhere [56-58]. In our sample, some subjects asked for clarification of whether activity limitation refers to a comparison with the performance of non-disabled subjects or the patient’s usual activities. Their reference point for “work or other regular daily activities” may be watching television. Similarly, “usual social activities” may mean getting a visit from a community nurse. This, combined with coarseness of the Role-Physical and Role-Emotional domain scales (which have ‘yes’ or ‘no’ responses in the Standard version [14] that we used) is likely to mean that the Role-Physical, Social Function and Role-Emotional scales underestimate the impact of spinal injury on these domain [59,60]. In our sample, these scales had large ceiling effects (54%, 44% and 77%, respectively) [30].

Tate et al [5] noted that people with spinal cord injuries ask, “What do you mean by health? I’m healthy and my health does not limit me but my spinal injury does.” They have also reported wide variations among patients in the extent to which they included the functional limitations resulting from their spinal cord injury in their concept of health [61]. This may reflect the ‘response shift’ [62] that can occur in chronic diseases, whereby some individuals adjust their internal standards, values and conceptualisations of health status in response to dramatic changes in their health and physical function, in order to achieve a psychological homeostasis. Unobtainable goals may be devalued over time and achievable goals become more important [63,64]. Thus, their perceptions and responses to the same question may change over time even though their apparent health does not. But response shift does not threaten the validity of self-report health status tools; rather, it complicates the interpretation of the resultant measures.

CONSTRUCT VALIDITY

Discriminative Validity

We have previously demonstrated [30] that this sample reported lower scores on the Physical Function, Role-Physical, Bodily Pain, General Health, Vitality and Social Functioning (but not Role-Emotional and Mental Health) domains and the Physical Composite scores than the Australian general population [22]. This demonstrates the SF-36’s can detect the gross impact of spinal cord injury.

The level of spinal cord injury has not consistently predicted health status [5], although many researchers [6,12,51] have found better physical health in paraplegics than tetraplegics. Our study confirmed the findings of Leduc and Lepage [12] that people with paraplegia had better Physical Function and Physical Composite scores, but not other SF-36 scores, when compared with people with tetraplegia. Unexpectedly, we also found that paraplegics reported more Bodily Pain ($p=0.01$) than tetraplegics and the difference in scores was small to moderate ($ES=0.31$). A similar association has been described by Andresen et al [8]. Interestingly, as there was no correlation between Bodily Pain and Physical Function domain scores in our sample, the individuals with good physical function were not necessarily those with more severe bodily pain.

Although complete injuries have been associated with poorer physical health [6], this association is not always evident [10,51]. The inability of our study to detect predicted differences between people with complete and incomplete spinal injuries on the Physical Function, Role-Physical and Physical Composite scales may reflect the definition of injury completeness that we used used, namely sensory *and* motor complete versus incomplete (i.e ASIA A versus B, C and D) [36]. It was not possible during our study to classify patients according to motor complete (ASIA A&B) versus incomplete (ASIA C&D). The latter distinction may better delineate differences in physical health which are likely to be linked most closely to motor rather than sensory disturbance.

In our sample, the SF-36 successfully detected the expected differences in health status between the employed and unemployed with spinal cord injury, although the differences were only small to moderate. These findings are consistent with most other studies [5,6,12,51,65]. Positive health status may facilitate employment, and employment may improve health status.

Although not universal [9], numerous authors have suggested worsening health status late in the lives of people with spinal cord injuries [5,12,65] and in the general population [22,26,66,67]. We found no association between current age and SF-36 scores. For the physical domain scores, this may be due in part to the overwhelming impact of the spinal injury relative to the more subtle effects of age-related comorbidities and impairments. Response shift may partially offset the expected effect of age on health status, as it is in general populations [13,22]. The ageing of persons with spinal injury has been raised as an area in need of research [68].

While many studies support the link between greater time since injury and better health status [6,9,51], others suggest that health status holds relatively steady throughout life [5,65] (but, as described above, health status may decline in the later years of life). Leduc and Lepage [12], who excluded persons less than two years after spinal injury, found no link between time since injury and SF-36 scores. Because of the inconsistency among studies and the opposing effects of increasing time since injury and aging, we were cautious in our predictions relating to time since injury. Despite the effect of ageing, in our sample those with the greatest time since injury had better Vitality, Social Function and Role-Emotional domain scores and Composite Mental scores (all predominantly psychological health measures), suggesting an adaptive process operating over a long time.

Previous studies suggest that while time since injury may be important for emotional well-being [6,9,51] after spinal cord injury, age at injury may be more important than current age in physical and emotional adjustment after spinal injury [30,51,69]. Older age at injury impedes recovery, perhaps related to a reduced ability to cope physically and cognitively, coupled with a lower vitality and a greater need for medical and social assistance [51,70].

In the general population, males tend to report better health status than females[22,26]: these differences tend to be small but are consistently observed. Leduc and Lepage [12] found that males with spinal cord injury had slightly higher Physical Function, Vitality and Mental Health scores than did females, while Westgren et al [51] confirmed this association only for Vitality scores. The gender differences in our sample were consistent in direction but were not statistically significant.

The SF-36 had greater discriminative power in the study by Leduc & Lepage [12]. In our study, 9 of 23 *a priori* hypotheses were confirmed (evidence ranging from borderline to strong statistical significance) and 9 of the remaining 14 differences were in the expected direction. Although our sample size was sufficient to detect clinically relevant differences, their study had more power (n=587 versus 305). Our sample may be less heterogeneous than Leduc and Lepage's [12], given our neurogenic bladder inclusion criterion. Sampling variation is also a possible explanation.

Most of the absolute differences in SF-36 scores between clinical and sociodemographic subgroups were not large in our sample. The SF-36 was primarily designed as a tool to reflect differences in the burden of disease between different populations and disease groups [71,72]. Among people within a disease group more subtle discrimination, for example by clinical and sociodemographic variables, might be better achieved with condition-specific measures of health status; in this case, a measure designed to assess issues relevant to people with spinal cord injury.

Convergent and Divergent Validity

The correlations between SF-36 domains in this sample were generally as observed in other populations [17], except for those involving the Physical Function scale, which were generally poor. This is likely to be due to the pronounced floor effects in the PF scale [30]. Given the content of the scales, the PF scale seems less prone to response shift than the remaining scales (which may explain why their inter-domain correlations are similar to those in non-disabled populations). The pattern of correlations we observed support Tate et al's [5,61] and Caplan's [73] observations that health status in people with spinal cord injury is more closely linked with participation restriction (better reflected in scales other than PF), than with physical impairment (which is best reflected in the PF scale [74]). This may explain the surprisingly tenuous link between physical impairment and health status in this population [5,8].

INTERNAL CONSISTENCY

The values for Cronbach's alpha in our sample, using the Standard version of the SF-36 [14], are similar to those in the Australian [24] and United States [14] general populations. Our results suggest that this version of the SF-36 has sufficient precision for use in population studies (e.g. quantifying burden of illness), but not for use in individual patient management (except for the Role-Physical and Role-Emotional scales) [48]. In Version 2 of the SF-36 [75], the Role-Physical and Role-Emotional domain scales have more item response options, which should increase their precision.

CONCLUSION

Our assessment of the SF-36's measurement properties supports its use in studies comparing groups with and without spinal injury, but suggest it may not be an ideal outcome measure for trials of interventions in people with spinal injuries or for use in individual patients. If used in clinical trials, it should be supplemented with condition-specific health status measures.

REFERENCES

1. Calman KC. Definitions and dimensions of quality of life. In: Aaronson NK, Beckman J, Eds. *The Quality of Life of Cancer Patients.* New York: Raven Press; 1987: 1-9.
2. Ware JE. Measuring functioning, well-being, and other generic health concepts. In: Osoba D, Ed. *Effect of Cancer on Quality of Life.* Boston: CRC Press; 1991.
3. Stockler MR, Osoba D, Corey P, Goodwin PJ, Tannock IF. Convergent Discriminative, and Predictive Validity of the Prostate Cancer Specific Quality of Life Instrument (PROSQOLI) – Assessment and Comparison with Analogous Scales from the EORTC QLQ-C30 and a Trial-specific Module. J Clin Epidemiol 1999; 52 (7): 653-66.
4. Gill TM, Feinstein AR. A critical appraisal of quality-of-life measurements. JAMA 1994; 272: 619-26.
5. Tate DG, Kalpakjian CZ, Forchheimer MB. Quality of life in individuals with spinal cord injury. Arch Phys Med Rehabil 2002; 83 Suppl 2: S18-25.
6. Lundqvist C, Siosteen A, Blomstrand C, Lind B, Sullivan M. Spinal cord injuries: Clinical functional, and emotional status. Spine 1991; 16 (1): 78-83.
7. Stensman R. Severely mobility-disabled people assess their quality of lives. Scand J Rehab Med 1985; 17: 87-99.
8. Andresen EM, Fouts BS, Romeis JC, Brownson CA. Performance of health-related quality of life instruments in a spinal cord injured population. Archives of Physical Medicine & Rehabilitation 1999; 80: 877-84.
9. Bach JR, Tilton MC. Life satisfaction and well-being measures in ventilator assisted individuals with traumatic tetraplegia. Arch Phys Med Rehabil 1994; 75: 626-32.
10. Craig AR, Hancock KM, Dickson H, Martin J, Chang E. Psychological consequences of spinal injury: A review of the literature. Australian and New Zealand Journal of Psychiatry 1990; 24: 418-25.
11. Craig AR, Hancock KM, Dickson H, Chang E. Long-term psychological outcomes in spinal cord injured persons: results of a controlled trial using cognitive behaviour therapy. Arch Phys Med Rehabil 1997; 78: 33-8.
12. Leduc BE, Lepage Y. Health-related quality of life after spinal cord injury. Disability and Rehabilitation 2002; 24(4): 196-202.
13. Jones PW. Measuring the quality of life of patients with respiratory diseases. Quality of Life: Assessment and Application.Edited by S Walker. London. MTP Press 1987. pp 235-251.
14. Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey - Manual & Interpretation Guide. The Health Institute, New England Medical Center, Boston, 1993.
15. Ware JE. The status of health assessment. Annu Rev Public Health 1995; 16: 327-54.
16. Ware JE, Davies AR. Monitoring health outcomes from the patients' point of view: A primer. Integrated Therapeutics Group, Inc., 1995.
17. Ware Jr JE, Gandek B; for the IQOLA Project. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. J Clin Epidemiol 1998; 51(11): 903-12.
18. Goller JL, McMahon JM, Rutledge C, Walker RG, Wood SE. Dialysis adequacy and self-reported health status in a group of CAPD patients. Advances in Peritoneal Dialysis 1997; 13: 128-33.
19. Adler DA, Bungay KM, Cynn DJ, Kosinski M. Patient-based health status assessments in an outpatient psychiatry setting. Psychiatric Services 2000; 51: 341-8.
20. Stewart AL, Hays Rd, Ware JE. Methods of validating the MOS health measures. In: Stewart AL, Ware JE, eds. Measuring functioning and wellbeing: the medical outcomes study approach. London: Duke University Press, 1992.
21. McHorney CA and Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research* 1995; 4: 293-307.
22. Australian Bureau of Statistics. 4399.0 National Health Survey: SF36 Population Norms, Australia. Canberra. Released 30/10/1997.
23. McCallum J. The SF-36 in an Australian sample: Validating a new, generic health status measure. Aust J Public Health 1995; 19 (2): 160-6.
24. Sanson-Fisher RW, Perkins JJ. Adaptation and validation of the SF-36 Health Survey for use in Australia. J Clin Epidemiol 1998; 51(11): 961-7.
25. Ward J, Lin M, Lajole V. Using MOS SF36 in general practice outcomes research. Proceedings of the Australian health outcomes and quality of life measurement conference; 1995 August 14-15, Canberra. Canberra: Australian Health Outcomes Clearing House, 1995: 68-74.
26. Brazier JE, Harper R, Jones NBM, et al. Validating the SF-36 health survey questionnaire: a new outcome measure for primary care. Br Med Journal 1992; 305: 160-4.

27. Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT. The SF-36 health survey questionnaire: an outcome measure suitable for use in the NHS? Br Med J 1993; 306: 1440-4.
28. Lyons RA, Perry HM, Littlepage BNC. Evidence for the validity of the short form 36 questionnaire (SF-36) in the elderly population. Age Ageing 1994; 23: 182-4.
29. McHornery CA, Ware JE, Rogers W, Raczek AE, et al. The validity and relative precision of the MOS short- and long-form health status scales and Dartmouth COOP charts: results from the medical outcomes study. Med Care 1992; 30 (suppl): MS 253-65.
30. Haran MJ, Lee BB, King MT, Marial O, Stockler M. Health status rated with the Medical Outcomes Study 36-Item Short-Form Health Survey after spinal cord injury. Arch Phys Med Rehabil. 86(12): 2290-5, 2005 Dec.
31. Lee B, Middleton JW, Rutkowski SB, Tangney B. The introduction of the NSW Spinal Injuries Database System. Proceedings of Paralympics Scientific Conference 2000, Sydney, Australia.
32. Yeo JD, Walsh J, Rutkowski S, Soden R, Craven M, Middleton J. Mortality following spinal cord injury. Spinal Cord 1998; 36: 329-336.
33. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Health Assessment Lab, New England Medical Centre, Boston, 1994.
34. Feinstein AR. Clinimetrics. New Haven, Connecticut: Yale University Press, 1987. pp 80.
35. McDowell I, Newell C. Measuring Health: A Guide To Rating Scales and Questionnaires. Second edition, Oxford University Press, New York 1996. Chapter 2: the theoretical and technical foundations of health measurement, pp30-46.
36. American Spinal Injury Association: International Standards for Neurological Classification of Spinal Cord Injury. Chicago. American Spinal Injury Association, 1992.
37. Sokal RR, Rohlf FJ. Biometry: The Principles And Practice Of Statistics In Biological Research. Third Edition. WH Freeman And Company, New York, 1995. pp423-39.
38. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 1988; 75 (4): 800-2.
39. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989; 27(Suppl): S178-89.
40. Cohen J. Statistical power analysis for the behavioural sciences. Second edition. Hillsdale NJ: Lawrence Erlbaum Associates; 1988.
41. Deyo RA, Diehr P, Patrick Dl. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation 1991; 12: 142S-158S.
42. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987; 40: 171-8.
43. Feinstein AR. Clinical epidemiology – the architecture of clinical research. Philadelphia: WB Saunders, 1985. pp170-90.
44. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951; 16: 297-334.
45. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. BMJ 1997; 314(7080): 572.
46. Helmstadter GC. Principles of psychological measurement. New York: Appleton-Century-Crofts, 1964.
47. Carmines E, Zeller R. Reliability and validity assessment. Quantitative applications in social science. Beverley Hills: Sage, 1979.
48. Nunnally JC. Psychometric Theory. New York: McGraw-Hill Publishing Company, 1978.
49. Minitab For Windows Software – Release 12. Copyright 1998. Minitab Inc. All rights reserved.
50. SPSS Statistical Software – Version 11.0. SPSS Inc., 2001. Chicago.
51. Westgren N, Levi R. Quality of life and traumatic spinal injury. Arch Phys Med Rehabil 1998; 79: 1433-9.
52. Devivo MJ, Rutt RD, Black KJ, Go BK, Stover SL. Trends in spinal cord injury demographics and treatment outcomes between 1973 and 1986. Arch Phys Med Rehabil 1992; 25: 86-91.
53. Treischman RB. Spinal Cord Injuries. New York: Pergammon Press, 1982.
54. Djikers M. Correlates of life satisfaction among persons with spinal cord injury. Arch Phys Med Rehabil 1999; 80: 867-76.
55. Kannisto M, Sintonen H. Later health-related quality of life in adults who have sustained spinal cord injury in childhood. Spinal Cord 1997; 35: 747-51.
56. Stadnyk K, Calder J, Rockwood K. Testing the measurement properties of the Short Form-36 Health Survey in a frail elderly population. J Clin Epidemiol 1998; 51(10): 827-35.
57. Kurtin PS, Ross Davies A, Meyer KB, DeGiacomo JM, Kantz ME. Patient-based status measures in outpatient dialysis: Early experiences in developing an outcomes assessment program. Medical Care 1992; 30 (5): MS136-49.
58. Mattson-Prince JM. A rational approach to long term care: comparing the independent living model with agency-based care for persons with high spinal cord injuries. Spinal cord 1997; 35: 326-31.

59. Kosinski M, Keller SD, Hatoum HT, et al. The SF-36 health survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis. Med Care 1999; 32 (5): MS10-22.
60. McHornery CA, Ware JE, Lu JFR, Sherbourne CD. The MOS-36 short form health survey (SF36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994; 32 (1): 40-66.
61. Tate D, Forchheimer M, Karunas R. Relationship of health status and functional independence to neurological impairment in spinal cord injury [abstract]. Arch Phys Med Rehabil 1998; 79: 1322.
62. Richards TA, Folkman S. Response Shift: a Coping Perspective. In Schwartz,C. E. and M. A. Sprangers. Adaptation to Changing Health: Response Shift in Quality-of-Life Research. Washington, American Psychological Association, 2000. pp 25-36.
63. Heckhausen J, Schultz R. A life-span theory of control. Psychol Rev 1995; 102: 284-304.
64. Myers D, Diener E. The pursuit of happiness. Sci Am 1996; 274(5): 70-2.
65. Krause JS, Crewe NM. Prediction of long-term survival of persons with spinal cord injury: an 11-year prospective study. Rehabil Psychol 1987; 32: 205-13.
66. Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Assessment of the SF-36 version 2 in the United Kingdom. J Epidemiol Community Health 1999; 53: 46-50.
67. Weinberger M, Samsa GP, Hanlon JT, et al . An evaluation of a brief health status measure in elderly veterans. J Am Geriatr Soc 1991; 39: 691-4.
68. Whiteneck GG. Changing attitudes toward life. Aging with spinal cord injury. New York: Demos Publications; 1993.
69. Widerstrom-Noga EG, Duncan R, Felipe-Cuervo E, Turk DC. Arch Phys Med Rehabil 2001; 83: 395-404.
70. Stensman R. Adjustment to traumatic spinal cord injury. A longitudinal study of self-reported quality of life. Paraplegia 1994; 32: 416-22.
71. Gatchell RJ, Polatin PB, Mayer TG, Robinson R, Dersh J. Use of the SF-36 Health Status Survey with a chronically disabled low back pain population: Strengths and limitations. J Occup Rehabil 1998; 8: 237-45.
72. Ware JEJ, Sherbourne CD. The MOS 36-item short form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30: 473-83.
73. Caplan B. Staff and patient perception of patient mood. Rehabil Psychol 1983; 28: 67-77.
74. State University of New York. Guide to the uniform dataset for medical rehabilitation (Adult FIM). Version 4.0. Buffalo. State University of New York at Buffalo, 1993.
75. Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Assessment of the SF-36 version 2 in the United Kingdom. J Epidemiol Community Health 1999; 53(1): 46-50.