Volume 11 • Number 4 • 2008
VALUE IN HEALTH

Validity, Responsiveness, and Minimal Important Difference for the SF-6D Health Utility Scale in a Spinal Cord Injured Population

Bonsan Bonne Lee, MBBS, FAFRM, MMED (Clin Epi), MHA, Gcert (IT),[1,2,3]
Madeleine T. King, BSc, GradDipMedStats, PhD,[4] Judy M. Simpson, PhD, CStat,[2]
Mark J. Haran, MBBS, FAFRM, MMed (Clin Epi), MHA,[5] Martin R. Stockler, MBBS MSc FRACP,[2]
Obaydullah Marial, MPH, Gdip (IT), Gcer (IT),[1] Glenn Salkeld, B.Bus, G Dip Health Economics, MPH, PhD[2]

[1]Prince of Wales Spinal Unit, Prince of Wales Hospital, Sydney, NSW, Australia; [2]University of Sydney, Sydney, NSW, Australia; [3]University of New South Wales, Sydney, NSW, Australia; [4]Centre for Health Economics Research and Evaluation, University of Technology, Sydney, NSW, Australia; [5]Royal North Shore Hospital, Sydney, NSW, Australia

ABSTRACT

Objective: To determine the feasibility, acceptability, discriminative validity, responsiveness, and minimal important difference (MID) of the SF-6D for people with spinal cord injury (SCI).

Methods: A total of 305 people with SCI completed the SF-36 health status questionnaire at baseline and at subsequent occurrence of a urinary tract infection (UTI) or 6-month follow-up. Normative SF-36 data were obtained from the Australian Bureau of Statistics. SF-36 scores were transformed to SF-6D utility values using Brazier's algorithm. We used UTI as the external criterion of clinically important change to determine responsiveness and two categories of the SF-36 transition question ("somewhat worse" and "somewhat better") as the external criterion to determine the MID. Derived SF-12 responsiveness was also assessed.

Results: The mean SF-6D values were: 0.68 (SD 0.21, n = 305) all patients; 0.66 (SD 0.19, n = 167) tetraplegia; 0.72 (SD 0.26, n = 138) paraplegia; 0.57 (SD 0.15, n = 138) with UTI. The Australian normative SF-6D mean value was 0.80 (SD 0.14, n = 18,005). The SF-6D was able to discriminate between SCI and the Australian normative sample (effect size [ES] = 0.86), tetraplegia–paraplegia (ES = 0.23), and it was responsive to UTI (ES = 0.86 SF-36 variant, ES = 0.92 SF-12 variant). The MID for respondents who reported being somewhat worse or somewhat better at follow-up was 0.03 (SD 0.17, n = 108/305), while the MID for only those who were somewhat worse was 0.10 (SD 0.14, n = 58).

Conclusions: The content of the SF-6D is more appropriate than that of the SF-36 for this physically impaired population. The SF-6D has discriminative power and is responsive to clinically important change because of UTI. The MID is consistent with published estimates for other disease groups.

Keywords: health economics, health preference state, SF-6D, spinal cord injury, urinary tract infection, validity.

Introduction

Preference-based measures of health allow the relative value of health states to be compared, both within and across diseases [1]. A fundamental concept underpinning this is health utility [2], a measure of preference for health outcomes. Combined with survival data, utilities can be used to estimate quality-adjusted life-years (QALYs). Utilities and QALYs are used in cost-utility analyses to assess the relative value of health interventions, across a range of purposes (preventive, diagnostic, curative, palliative), types (programs, services, technologies, pharmaceuticals), and populations (within and across diseased, disabled, and healthy populations). Preference-based measures are therefore useful and important outcome measures for policymakers, both locally and internationally.

The SF-6D, a relatively new utility measure, is particularly attractive as it is calculated from the SF-36, a health status measure commonly used to assess the impact of disease and disability, including spinal injury [3]. In common with other multiattribute utility instruments such as EQ-5D (Euroqol) and Health Utilities Index (HUI) [4], it allows those experiencing the health states to contribute directly to utility scores. A particular advantage of the SF-36 and SF-6D is that they economize on data collection, yielding measures of both health status and utility. Since the SF-6D methodology was published 5 years ago, it has rapidly become a popular method of utility estimation. A recent systematic review of the use of heath status measurement instruments to calculate QALYs found

Address correspondence to: Bonsan Bonne Lee, PO Box 431, Broadway, Sydney, NSW 2007, Australia. E-mail: blee@medicalinformatics.net
10.1111/j.1524-4733.2007.00311.x

  1098-3015/08/680 680–688

that, despite its contemporary origin, the SF-6D accounted for 5% of the instruments used [5].

The SF-6D is a utility measure based on a six-dimensional health state classification. It is derived from a subset of 11 SF-36 questions covering the dimensions of Physical Functioning, Role Limitation, Social Functioning, Pain, Mental Health, and Vitality. It allows a possible 18,000 health states to be defined. A survey (involving SF-6D, ranking, and standard gamble) of 249 health states defined by the SF-6D was valued by a representative sample of the UK general public (n = 611). Econometric methods were then used to determine a model for predicting the standard gamble scores generated by the valuation survey [6]. Brazier et al. have shown that the SF-6D is a viable alternative preference measure [6]. It can be derived from either the SF-36 [6] or the shorter SF-12 [7]. It has been suggested that the SF-6D may be more sensitive than the EQ-5D, especially for mild–moderate health issues [6]. Limitations and outstanding issues with the SF-6D include whether it compromises the richness of the original SF-36 [6] and whether it is less sensitive when used in poorer health states [6,8,9]. It is therefore important that additional validation studies are performed in different populations and settings. The present article describes such a validation in an Australian population with spinal cord injury (SCI), most of whom were living in a general community setting.

The minimal important difference (MID) allows clinicians to determine whether a change observed on a self-reported health rating scale is meaningful or trivial. It has been defined as the smallest difference in score that the patient perceives as beneficial [10]. For our purposes, in the absence of significant side effects or cost barriers, this would lead to a change in clinical decision-making.

This article provides the first validation and MID values for the SF-6D in the SCI population. We assess the acceptability and appropriateness of the SF-6D for application in SCI, evaluate its discriminative ability, and determine its responsiveness to clinically important change. The external criterion used to define clinically important change is the occurrence of a urinary tract infection (UTI), a common comorbidity in this patient population group, with a reported incidence of 1.82 episodes per annum [11,12].

Methods

Data were collected during the Spinal-Injured Neuropathic Bladder Antisepsis randomized controlled trial [13]. Subjects were sampled from the New South Wales (State) Spinal Cord Injuries Database [14] and related databases of two major teaching hospitals. Inclusion criteria were: SCI with neurogenic bladder; stable bladder management; absence of serious renal pathology; not taking antibiotics at enrollment; and absence of symptoms of a UTI at enrollment. Between November 2000 and August 2002, 543 eligible patients (predominantly community dwelling) were invited to participate in the study, of whom 305 (56%) agreed. Characteristics of the sample and reasons for nonparticipation are reported elsewhere [13].

Subjects completed the SF-36 at enrollment and again on development of UTI. If no UTI was experienced, a repeat SF-36 was completed at 6-month follow-up. Subjects completed the SF-36 by self-report with a research officer present, or by self-report via mail. Incomplete responses or inconsistencies were clarified by direct inquiry. Interpreters and physical assistance were used where necessary.

The SF-6D utility and dimensional scores were derived from SF-36 responses using Brazier's algorithm [6]. The domains and SF-36 items [15] used to construct the SF-6D were: Physical Functioning (items 3a, 3b, and 3j); Role Limitation due to physical problems (item 4c) and Role Limitation due to emotional problems (item 5b); Social Functioning (item 10); both bodily pain items (items 7 and 8); Mental Health (items 9b and 9f); and Vitality (item 9e). To explore whether the SF-12 version of the SF-6D differed in responsiveness from the full SF-36 version, SF-6D utility scores were recalculated using the Brazier SF-12 algorithm [7] for the responsiveness analysis.

Acceptability and appropriateness were assessed in terms of feasibility and content validity. Practical difficulties in the use of the SF-36 or content issues identified by subjects, research assistants, or authors during administration were recorded. Ceiling and floor effects for each SF-6D domain were examined by neurological level of injury (tetraplegia vs. paraplegia) [16].

Discriminative validity was assessed with cross-sectional comparisons of mean SF-6D utility and dimensional scores, externally by comparing the SCI patients with Australian normative data, and internally by comparing various subgroups of the SCI patients. Normative Australian SF-36 data were from 18,005 respondents in the Australian Bureau of Statistics National Health Survey of 1995 [17]. Means and standard deviations for normative and sample data were adjusted to fit the age and sex distribution of the Australian population using direct standardization [18].

Internal cross-sectional comparisons were based on a priori hypotheses made by three experts, one in rehabilitation medicine (BL) and two in quality-of-life research (MK, MS), who independently ranked their expectations about the size and direction of differences in SF-36 scales between groups defined by six clinically relevant variables. When applied to the single SF-6D index score, these led to six a priori hypotheses about the derived utility scores: that more (vs. less) extensive neurological level, more (vs. less) completeness of