Association of Academic Surgery

A comparison of the validity of two indirect utility instruments as measures of postoperative recovery

CrossMark

Lawrence Lee, MD, MSc,[a,*] Juan Mata, MD,[a] Berson R. Augustin,[b]
Franco Carli, MD,[b] Nancy Morin, MD,[c] Eric Latimer, PhD,[d]
and Liane S. Feldman, MD[a]

[a] Department of Surgery, Steinberg-Bernstein Centre for Minimally Invasive Surgery and Innovation, Montreal, Quebec
[b] Department of Anaesthesia, McGill University Health Centre, Montreal, Quebec
[c] Department of Surgery, Sir Mortimer B. Davis Jewish General Hospital, Montreal, Quebec
[d] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec

ARTICLE INFO

ABSTRACT

Background: Cost-effectiveness analyses of surgical interventions require valid measures of postoperative recovery. The objective of this study was to compare the validity of two indirect utility instruments, the Short Form 6D (SF-6D) and EuroQol 5D (EQ-5D), as measures of postoperative recovery.

Materials and methods: A prospective cohort of patients undergoing elective colorectal resection at two university-affiliated institutions from October 2012–October 2013 completed the SF-6D and EQ-5D (including the EQ-visual analog scale [EQ-VAS]) at baseline (before surgery), and at 4 and 8 wk after surgery. Responsiveness and construct validity were assessed through *a priori* hypotheses.

Results: A total of 165 patients were included. The SF-6D was the most responsive to the expected postoperative changes at 4 and 8 wk compared with the EQ-5D and the EQ-VAS. The 4-wk SF-6D, EQ-5D, and EQ-VAS discriminated between patients with and without complications after controlling for confounders with adjusted mean differences of −0.070 (95% confidence interval [CI] −0.126 to −0.015), −0.133 (95% CI −0.231, −0.030), and −7.91 (95% CI −14.77, −1.04), respectively. Mean SF-6D and EQ-5D values were significantly different from the US population norms at all time points, but the magnitude of change was highest for the SF-6D. The strength of correlation between all three instruments was moderate at all time points ($r = 0.550$–0.684, all $P < 0.05$).

Conclusions: The SF-6D preference-based health index appears to be a more valid measure of postoperative recovery than the EQ-5D and EQ-VAS in surgical cost-effectiveness analyses.

© 2014 Elsevier Inc. All rights reserved.

1.   Introduction

Many new surgical innovations are advocated to improve recovery, but few have been fully evaluated. Cost-effectiveness analyses of these new technologies require valid measures of postoperative recovery. However, it is difficult to compare studies with one another as multiple different outcomes have been used to measure "recovery" [1–3]. The Panel on Cost-Effectiveness in Health and Medicine recommends using quality-adjusted life years (QALYs) to measure effectiveness in economic evaluations, as QALYs incorporate multiple different outcomes into a single summary measure [4]. QALYs are calculated by multiplying the duration spent in a particular health state by the quality-of-life weight of that health state, which is measured in terms of "utilities." Utilities are measured on a scale from 0.0 (death) to 1.0 (perfect health). Although QALYs can be directly elicited, they are commonly measured using indirect utility instruments, such as the Short Form 6D (SF-6D) [5] or the EuroQol 5D (EQ-5D) [6].

However, there may be significant variability in utility values depending on the instrument used, which can have a large effect on the interpretation of cost-effectiveness results [7]. The choice of which utility instrument to use depends on an instrument's practicality, reliability, and validity for the specific condition under investigation [8]. For conditions such as knee pain, the EQ-5D is superior to the SF-6D [9], but the opposite is true for rheumatoid arthritis [10]. The SF-6D has been previously validated as a measure of postoperative recovery after colorectal surgery specifically to use in cost-effectiveness analyses [11], but the EQ-5D is more commonly used and easier to administer [12]. No previous study has investigated the validity of the EQ-5D to value postoperative recovery. Therefore, the objective of this study was to compare the validity of two important indirect utility instruments, the SF-6D and EQ-5D, as measures of postoperative recovery in patients undergoing elective colorectal resection.

2.   Materials and methods

2.1.   Patients

Consecutive patients undergoing elective colorectal surgery at two university-affiliated institutions between October 2012 and October 2013 were approached for participation at the preoperative clinic (1–2 wk before surgery). Patients were eligible if they were aged >18 y and had a scheduled resection of the colon and/or rectum. Patients were excluded if they did not speak English or French or had neuropsychiatric conditions or cognitive impairments that interfered with completion of the study questionnaires. Patients who did not undergo resection during the operation were also excluded. Demographics, operative details, postoperative course, and final pathology were collected prospectively. Comorbidities were measured using the Charlson Comorbidity Index [13]. Postoperative complications were graded as per the Clavien classification of surgical complications [14]. Outcomes were collected up to 60 d postoperatively.

2.2.   Measures

All participating patients completed the SF-36 [15], from which the SF-6D is derived, and the EQ-5D [16] at baseline (within 2 wk before surgery) and at 4 and 8 wk after surgery. The SF-6D is a multiattribute indirect utility measure that is nested within the SF-36 [5]. It measures six dimensions: physical functioning, role limitations, social functioning, pain, mental health, and vitality, for a possible 18,000 unique health states. Quality of life weights were obtained from a representative sample of 611 members, the UK general population using standard gamble [5]. The SF-6D ranges from 0.296 (the most severe impairment in all six dimensions) to 1.000 (perfect health).

The EQ-5D is a five-item instrument, that is, measuring mobility, self-care, usual activities, pain, and anxiety or depression. Each item is scored on a three-level scale for a possible 243 unique health states. Different valuations have been published, but the UK tariff [6] (derived from a representative sample of 3395 members of the UK general population using time trade-off) was used to maintain comparability with the SF-6D. The EQ-5D ranges from −0.594 (the most severe impairments in all five dimensions) to 1.000 (perfect health). The EQ-5D also contains a visual analog scale (EQ-VAS), which asks respondents to rate their present health on a scale from 0 (the worst imaginable health state) to 100 (the best imaginable health state).

2.3.   Validity

All three measures were examined for responsiveness and construct validity. All hypotheses were made *a priori* and were based on previous validity data [11]. In order to be considered responsive, we hypothesized that each measure would be lower at 4 wk compared with baseline, improve from 4–8 wk and return to baseline by 8 wk. This analysis was also stratified by resection type (colonic *versus* rectal) to account for differences in postoperative functional outcomes (which in turn are indirectly measured by these instruments). Construct validity was assessed in several ways: first, we hypothesized that each measure would discriminate between patients with and without complications at 4 wk, between colonic and rectal resections at all time points and between patients receiving adjuvant therapy by 8 wk and those who did not; and second, each measure would be different from population norms at all time points. The US population mean values for the SF-6D, EQ-5D (using the UK tariff), and EQ-VAS were obtained from a study by Hanmer *et al.* [17].

2.4.   Statistical analysis

Summary descriptive data were expressed as proportion (*n*), mean (standard deviation [SD]), or median interquartile range (IQR), as appropriate. Correlations between each measure were demonstrated using Pearson correlation (*r*). Bland–Altman plots were used to demonstrate the agreement between measures [18]. In these plots, the difference in scores is plotted against the mean score, along with the limit of agreement (±2 SD). If the limits of agreement between scores are not clinically