



Published in final edited form as:

Spine (Phila Pa 1976). 2009 September 1; 34(19): 2085–2089. doi:10.1097/BRS.0b013e3181a93ea6.

Predicting SF-6D utility scores from the Oswestry Disability Index and Numeric Rating Scales for Back and Leg Pain

Leah Y. Carreon, MD, MSc¹, Steven D. Glassman, MD¹, Christine M. McDonough, PhD, PT, MS^{2,3}, Raja Rampersaud, MD⁴, Sigurd Berven, MD⁵, and Michael Shainline, MS, MBA⁶

¹Kenton D. Leatherman Spine Center, 210 East Gray Street, Suite 900, Louisville, KY, USA

²The Health and Disability Research Institute at Boston University, 580 Harrison Avenue 2ND Floor, Boston, MA 02118

³The Dartmouth Institute for Health Policy and Clinical Practice, 565 Ruben Clinical Research Section, Dartmouth-Hitchcock Medical Center, One Medical Center Drive, Lebanon, NH 03756

⁴Department of Surgery, Division of Orthopaedic and Neurosurgery, University of Toronto, Toronto, ON, Canada

⁵Department of Orthopaedic Surgery, University of California San Francisco, San Francisco, CA, USA

⁶Biostatistics, PhDx Systems, Inc., Albuquerque, NM, USA

Abstract

Study Design—Cross-sectional cohort

Objective—The purpose of this study is to provide a model to allow estimation of utility from the SF-6D using data from the ODI, BPNRS, and the LPNRS.

Summary of Background Data—Cost-utility analysis provides important information about the relative value of interventions and requires a measure of utility not often available from clinical trial data. The Oswestry Disability Index (ODI) and numeric rating scales for back (BPNRS) and leg pain (LPNRS), are widely used disease-specific measures for health-related quality of life in patients with lumbar degenerative disorders. The purpose of this study is to provide a model to allow estimation of utility from the SF-6D using data from the ODI, BPNRS, and the LPNRS.

Methods—SF-36, ODI, BPNRS and LPNRS were prospectively collected pre-operatively, at 12 and 24 months post-operatively in 2640 patients undergoing lumbar fusion for degenerative disorders. Spearman correlation coefficients for paired observations from multiple time points between ODI, BPNRS and LPNRS and SF-6D utility scores were determined. Regression modeling was done to compute the SF-6D score from the ODI, BPNRS and LPNRS. Using a separate, independent dataset of 2174 patients in which actual SF-6D and ODI scores were available, the SF-6D was estimated for each subject and compared to their actual SF-6D.

Results—In the development sample, the mean age was 52.5 ± 15 years and 34% were male. In the validation sample the mean age was 52.9 ± 14.2 years and 44% were male. Correlations between the SF-6D and the ODI, BPNRS and LPNRS were statistically significant ($p < 0.0001$) with correlation coefficients of 0.82, 0.78, and 0.72 respectively. The regression equation using ODI, BPNRS and LPNRS to predict SF-6D had an R^2 of 0.69 and a root mean square error

(RMSE) of 0.076. The model using ODI alone had an R^2 of 0.67 and a RMSE of 0.078. The correlation coefficient between the observed and estimated SF-6D score was 0.80. In the validation analysis, there was no statistically significant difference ($p=0.11$) between actual mean SF-6D (0.55 ± 0.12) and the estimated mean SF-6D score (0.55 ± 0.10) using the ODI regression model.

Conclusion—This regression-based algorithm may be used to predict SF-6D scores in studies of lumbar degenerative disease that have collected ODI but not utility scores.

Keywords

SF-6D; Oswestry Disability Index; cost-utility analysis; utilities; outcome measurement

INTRODUCTION

Cost-utility analysis is increasingly used by decision makers to assess the relative value of alternative treatment interventions in the context of limited resources [1,2] In order to calculate Quality-Adjusted Life-Years (QALYs) for cost-utility analysis, outcomes of treatment are measured using a single score, anchored at 0 for death and 1 for perfect health, and weighted for the relative desirability of the health state. Standards for economic evaluations recommend using societal values (utilities or preferences) [3]. There are two main approaches to obtaining “societal health state values”: 1) direct measurement of value for health states of a representative sample of the population using methods such as standard gamble, time tradeoff, and visual analogue scale ratings, and 2) indirect measurement using preference-based measurement systems such as the Quality of Well Being Scale [4], the EuroQOL EQ-5D [5], SF-6D [6] or the Health Utilities Index HUI [7]. The SF-6D uses a subset of items from the widely used SF-36 as the basis of its health state classification system.

Standards for measurement of outcomes in clinical research for spine disorders include recommendations for the SF-12 or the EQ-5D to measure general health and to allow cost-effectiveness analysis [8] Serious tradeoffs involved in choosing outcome measures for clinical trials, including the need to obtain adequate power to detect differences and to maximize measurement precision, must be balanced against practicality and limits in available resources. With these concerns in mind, unless cost-utility analysis is a specific research aim, investigators often favor disease-specific measures that appear to focus on the key aspects targeted by treatment over generic preference-based instruments. In reality, many clinical trials in spine disorders utilize disease-specific measures including the Oswestry Disability Index (ODI) and numeric rating scales for back and leg pain instead of preference-based instruments.

In order to allow cost-utility analysis using available clinical trial data, several authors have developed various methods to predict or “map” societal health state values using data from non-preference-based instruments [9–29]. The purpose of this study is to determine whether the widely used disease-specific measures (ODI, numeric rating scales for back and leg pain) may accurately predict SF-6D utility scores.

METHODS

The Development Sample

As part of a multi-center database, the Low Back 2000, the Medical Outcome Study Short Form-36 (SF-36) [30], the Oswestry Disability Index (ODI), Back Pain Numeric Rating Scale (BPNRS, 0 to 10) and Leg Pain Numeric Rating Scale (LPNRS, 0 to 10) were prospectively collected pre-operatively, at 12 and 24 months post-operatively in 2640

patients undergoing lumbar fusion for degenerative disorders from January 2002 to June 2004.

The Validation Sample

Patients undergoing lumbar fusion for degenerative disorders from September 2003 to December 2005 from a second separate multi-center database, the Low Back 2004, were used as a validation sample. The same outcome measures, the SF-36, ODI, BPNRS and LPNRS, were prospectively collected at similar time points, pre-operatively, at 12 and 24 months post-operatively.

Outcome Measures

ODI—The Oswestry Disability Index [31] is a self-administered questionnaire measuring “back-specific function” on a 10 item scale with six response categories each. Each item scores from 0 to 5, higher scores being worse, which is transformed into a 0–100 scale. The ten items include pain intensity, personal care, lifting, walking, sitting, standing, sleeping, work, social life and traveling. Patients with scores between 0 to 20 have Minimal Disability, between 21 and 40 have Moderate Disability, between 41–60 have Severe Disability, 61 to 80 are crippled and 81 to 100 are bed-bound or exaggerating their symptoms.

SF-6D—The SF-6D is a preference-based health state classification system derived from the SF-36 that defines 18,000 health states [32]. It was constructed using a selection of 11 items from the SF-36 across six of the eight SF-36 dimensions of health; physical functioning, role limitation, social functioning, bodily pain, mental health and vitality. There are 3 states for physical functioning, 4 for role limitations and 5 each for social functioning, bodily pain, mental health and vitality, producing 18,000 different health states. A subset of 249 SF-6D states were valued by a representative sample of the UK general population (n=611) using the standard gamble valuation technique. Regression models were used to estimate health state values for the full range of SF-6D health states. The resultant algorithm can be used to convert SF-36 data at the individual level to societal health state values or preference scores.

BPNRS/LPNRS—The Numeric Rating Scale are two items, one each for back pain and leg pain, on the survey that asks: “On a scale from 0 to 10, mark your level of back (leg) pain discomfort, with 0 being none and 10 being unbearable.”[33]

Statistical Analysis—All correlation analyses were between the SF-6D and the corresponding paired ODI score or BPNRS or LPNRS. Spearman and Pearson coefficients for paired observations from multiple time points between SF-6D and ODI were determined. Although these measures can be analyzed as continuous variables, some argue that these measures are long ordinal. Spearman correlation coefficients between SF-6D and BPNRS and LPNRS were calculated as the BPNRS and LPNRS are considered discrete ordinal scales. Linear Regression modeling was conducted to predict SF-6D scores from the ODI, BPNRS and LPNRS scores. The R^2 and root mean square error (RMSE) estimate was used to assess the ability of the model to predict SF-6D scores, and to compare the performance of the various models.

To evaluate external validity of the final regression model, the algorithm generated using the Low Back 2000, was applied to paired observations of SF-6D, ODI and BPNRS and LPNRS from a separate, independent database of patients undergoing lumbar fusion surgery for degenerative disorders (Low Back 2004). The SF-6D was estimated for each subject and

compared to their actual SF-6D in this independent dataset. Health state values for each individual paired observation were compared using paired t-tests.

RESULTS

There were 2640 patients in the development sample, and 2174 in the validation sample. In the development sample, the mean age was 52.2 ± 15.0 years and 44.8% were male. In the validation sample the mean age was 52.9 ± 14.2 years and 44.2% were male. There was no statistically significant difference in the age and gender distribution between the two samples. Tables 1 and 2 provide a summary of the outcome measures for the two samples. Correlations between the SF-6D and the ODI, BPNRS and LPNRS were statistically significant ($p < 0.0001$). The strongest correlation was for SF-6D and ODI (Pearson = 0.83, Spearman = 0.82). This was followed by SF-6D and BPNRS (0.78). The relationship between SF-6D and LPNRS was the weakest (0.72) but it was still statistically significant. Mean scores for SF-6D, BPNRS and LPNRS across the different Oswestry Disability categories is shown in Table 3.

The regression equation using ODI, BPNRS and LPNRS to predict SF-6D had a correlation coefficient of 0.83 and accounted for 69% of the variability with an RMSE of 0.076 (Table 4). The regression equation is

$$\text{SF-6D} = 0.79924 - 0.00409(\text{ODI}) - 0.00783(\text{BPNRS}) - 0.00296(\text{LPNRS})$$

Using only the ODI in the regression equation:

$$\text{SF-6D} = 0.78275 - 0.00518(\text{ODI})$$

decreased the correlation coefficient minimally to 0.82 and still accounted for 67% of the variability of SF-6D with an RMSE of 0.078. This linear relationship can be seen in the SF-6D/ODI plot (Figure 1).

The results of the confirmatory analysis showed that the estimated SF-6D using the regression equation with ODI alone was very similar to the actual SF-6D. The correlation coefficient between the observed and estimated SF-6D was 0.80. There was no statistically significant difference ($p = 0.110$) between the actual SF-6D (0.55 ± 0.12) and the estimated SF-6D (0.55 ± 0.10) using the ODI regression model in the Low Back 2004 dataset (Figure 2).

DISCUSSION

Measurement of health utility states is important for clinical outcomes research to permit cost-utility analyses, and to compare efficacy of healthcare interventions across different areas. Utility states are required for the calculation of QALYs, and this permits a unit to be assigned to the value of a healthcare intervention. In spine surgery, most clinical outcome studies have emphasized disease-specific measures of health-related quality of life. A technique to use standard disease-specific measures to estimate utility scores is valuable because it would permit the use of existing data to determine cost-utility and QALY values. This study of 2640 patients with lumbar degenerative disorders provides a regression model to predict SF-6D utility scores using the most widely used disease-specific patient reported outcome in low back pain, the ODI. Our results suggest that the relationship between the SF-6D and the ODI is sufficiently robust to allow a valid estimation of SF-6D scores from

the ODI using a regression equation when health state values were not measured using a preference-based instrument. This finding is consistent with comparable clinical responsiveness shown between ODI and SF-36 in the spine population [34, 35]. The pain and function subscales of the SF-36, which account for 7 of the 11 questions used for calculation of the SF-6D, have demonstrated equal sensitivity to change compared to ODI. The performance of the model in predicting the actual mean utility scores in a large, independent validation sample further supports the use of this method to estimate SF-6D utility scores. Although our model performed well in predicting SF-6D utility scores as measured by the RMSE, the ceiling effect for predicted health state values should be noted. The implication of this characteristic of our model is a limited ability to detect change at the high end of health status. However, the R^2 of 0.67 for our model is superior to the entire range of models similarly estimating health state values from disease-specific measures reported by Mortimer and Segal [36]. Ideally, health state values would be measured in clinical studies using a preference-based measurement system. However, there are many circumstances in which such data are not collected at the time of the clinical study. Providing a method to estimate societal health state values using data from the ODI may facilitate the conduct of cost-utility analyses of interventions that were previously unachievable. Using the ODI as a predictor of health state may also reduce respondent and administrative burden, particularly in an institutional or clinical setting, where the use of the simple, one-page, 10-item ODI may be ideal.

This significant contribution must be weighed against limitations of this approach for estimating health state values. Some studies indicate that estimates from mapping may have important limitations, including limited variability and measurement precision. [4, 7, 27] A practical implication of these disadvantages for clinical researchers may be a diminished ability to detect significant differences in health outcomes. Furthermore, studies have shown that different approaches to converting health status to health state values may produce meaningfully different estimates [4, 7, 27]. A recent review of approaches for converting health status into health state values recommended that researchers consider the validity and feasibility of the method in the disease area over a particular technique [36] The ODI is widely used and validated and the score is determined using a specific questionnaire and scoring algorithm [31, 37]. There are, however, modified, unvalidated versions of the questionnaire. The SF-6D and ODI measure several similar domains that may be affected by low back pain. The “Social Function” domain in the SF-6D and “Social Life” item in the ODI both measure the effect of low back pain on the individual’s social life; the “Role limitation” domain in the SF-6D and the “Personal Care” and “Employment/Homemaking” items in the ODI address the impact of low back pain on the patient’s work and daily activities.

In contrast both the BPNRS and LPNRS address only the intensity of the pain in the patient’s back or leg and not how the disease affects the patient’s quality of life. This may be the reason why these two measures do not add much to the regression model. The method by which back and leg pain are measured vary widely from study to study in terms of value ranges used, and may include both intensity and frequency of the pain and not just the intensity of pain. The fact that BPNRS and LPNRS do not add much to the model allows for increased applicability.

Using the equation to estimate SF-6D in an independent dataset showed that the estimated SF-6D using the ODI regression model was very similar to the actual SF-6D. There was no statistically significant difference between the actual SF-6D and the estimated SF-6D using the ODI regression model. As there is a strong correlation between ODI and the SF-6D, which is consistent across time periods and databases examined in this study, the ODI

regression model can be used to estimate SF-6D utility scores in studies of lumbar degenerative disease that have collected ODI but not utility scores.

Acknowledgments

The authors would like to acknowledge the support of the Lumbar Spine Study Group.

References

1. Bloom BS. Use of formal benefit/cost evaluations in health system decision making. *Am J Manag Care.* 2004 May; 10(5):329–35. [PubMed: 15152703]
2. Dickson, M.; Hurst, J.; Jacobzone, S. Survey of pharmacoeconomic assessment activity in eleven countries. OECD; 2003.
3. Gold M, Franks P, et al. Assessing the health of the nation. The predictive validity of a preference-based measure and self-rated health. *Medical Care.* 1996; 34(2):163–77. [PubMed: 8632690]
4. Kaplan RM. Health-related quality of life in cardiovascular disease. *J Consult Clin Psychol.* 1988 Jun; 56(3):382–92. [PubMed: 3294264]
5. The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy.* 1990; 16:199–208. [PubMed: 10109801]
6. 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. [PubMed: 11939242]
7. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single attribute utility functions for the health utilities index mark 3 system. *Med Care.* 2002; 40:113–28. [PubMed: 11802084]
8. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G. Outcome measures for low back pain research. A proposal for standardized use. *Spine.* 1998 Sep 15; 23(18):2003–13. [PubMed: 9779535]
9. Buxton, MJ.; Lacey, LA.; Feagan, BG.; Niecko, T.; Miller, DW.; Townsend, RJ. Mapping from Disease-Specific Measures to Utility: An Analysis of the Relationships between the Inflammatory Bowel Disease Questionnaire and Crohn's Disease Activity Index in Crohn's Disease and Measures of Utility.
10. Brazier J, Kolotkin RL, Crosby RD, et al. Estimating a preference-based single index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) instrument from the SF-6D. *Value in Health.* 2004; 7:490–8. [PubMed: 15449641]
11. Erickson P. Evaluation of a population-based measure of quality of life: the Health and Activity Limitation Index (HALex). *Quality of Life Research.* 1998; 7:101–14. [PubMed: 9523491]
12. Franks P, Lubetkin EI, Gold MR, et al. Mapping the SF-12 to preference-based instruments. *Medical Care.* 2003; 41:1277–83. [PubMed: 14583690]
13. Fryback DG, Lawrence WF, Martin PA, et al. Predicting Quality of Well-being scores from the SF-36: results from the Beaver Dam Health Outcomes Study. *Medical Decision Making.* 1997; 17:1–9. [PubMed: 8994146]
14. Gold M, Franks P, Erickson P. Assessing the health of the nation. The predictive validity of a preference-based measure and self-rated health. *Medical Care.* 1996; 34:163–77. [PubMed: 8632690]
15. Gold MR, Franks P, McCoy KI, et al. Toward consistency in cost-utility analyses: using national measures to create condition-specific values. [see comment]. *Medical Care.* 1998; 36:778–92. [PubMed: 9630120]
16. Gray AM, Rivero-Arias O, Clarke PM. Estimating the Association between SF-12 Responses and EQ-5D Utility Values by Response Mapping. *Med Decis Making.* 2006; 26:18–29. [PubMed: 16495197]
17. Hollingworth W, Deyo RA, Sullivan SD, et al. The practicality and validity of directly elicited and SF-36 derived health state preferences in patients with low back pain. *Health Economics.* 2002; 11:71–85. [PubMed: 11788983]
18. Kind P, Macran S. Eliciting social preference weights for functional assessment of cancer therapy-lung health states. *Pharmacoeconomics.* 2005; 23:1143–53. [PubMed: 16277549]

19. Lawrence WF, Fleishman JA. Predicting EuroQoL EQ-5D preference scores from the SF-12 Health Survey in a nationally representative sample. *Medical Decision Making*. 2004; 24:160–9. [PubMed: 15090102]
20. Lee TA, Hollingworth W, Sullivan SD. Comparison of directly elicited preferences to preferences derived from the SF-36 in adults with asthma. *Medical Decision Making*. 2003; 23:323–34. [PubMed: 12926582]
21. Lenert LA, Sherbourne CD, Sugar C, et al. Estimation of utilities for the effects of depression from the SF-12. *Medical Care*. 2000; 38:763–70. [PubMed: 10901359]
22. Lenert LA, Sturley AP, Rapaport MH, et al. Public preferences for health states with schizophrenia and a mapping function to estimate utilities from positive and negative symptom scale scores. *Schizophrenia Research*. 2004; 71:155–65. [PubMed: 15374583]
23. Longworth L, Buxton M, Sculpher M, Smith DH. Estimating utility data from clinical indicators for patients with stable angina. *Eur J Health Econ*. 2005; 6:347–53. [PubMed: 16193322]
24. Lundberg I, Johannesson M, Isacson D, et al. The relationship between health state utilities and the SF-12 in a general population. *Med Decis Making*. 1999; 12:128–40. [PubMed: 10231075]
25. Sanderson K, Andrews G, Corry J, et al. Using the effect size to model change in preference values from descriptive health status. *Quality of Life Research*. 2004; 13:1255–64. [PubMed: 15473504]
26. Sengupta N, Nichol MB, Wu J, Globe D. Mapping the SF-12 to the HUI3 and VAS in a managed care population. *Med Care*. 2004; 42:927–37. [PubMed: 15319619]
27. Sherbourne C, Unutzer J, Schoenbaum M, et al. Can utility-weighted health-related quality-of-life estimates capture health effects of quality improvement for depression? *Medical Care*. 2001; 39:1246–59. [PubMed: 11606878]
28. Shmueli A. Subjective health status and health values in the general population. *Medical Decision Making*. 1999; 19:122–7. [PubMed: 10231074]
29. Sullivan PW, Ghushchyan V. Mapping the EQ-5D Index from the SF-12: US general population preferences in a nationally representative sample. *Med Decis Making*. 2006; 26:401–9. [PubMed: 16855128]
30. Ware, JE.; Kosinski, M.; Keller, SK. SF-36 Physical and mental health summaries scales: A user's manual. Boston, MA: The Health Institute; 1994.
31. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry Low Back Pain Questionnaire. *Physiotherapy* Aug. 1980; 66(8):271–273.
32. Brazier J, Usherwood T, Harper R, Thomas D. Deriving a preference-based single index from the UK SF36 Health Survey. *J Clin Epidemiol*. 1998; 51(11):1115–1128. [PubMed: 9817129]
33. Jensen MP, Turner JA, Romano JM. Correlates of improvement in multidisciplinary treatment of chronic pain. *Journal of Consulting and Clinical Psychology*. 1994; 62:172–179. [PubMed: 8034820]
34. Taylor SJ, Taylor AE, Foy MA, Fogg AJ. Responsiveness of common outcome measures for patients with low back pain. *Spine*. 1999 Sep 1; 24(17):1805–12. [PubMed: 10488511]
35. Walsh TL, Hanscom B, Lurie JD, Weinstein JN. Is a condition specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. *Spine*. 2003; 28:607–15. [PubMed: 12642770]
36. Mortimer D, Segal L. Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY weights. *Medical Decision Making*. 2008; 28:66–89. [PubMed: 18263562]
37. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther*. 2001 Feb; 81(2):776–88. [PubMed: 11175676]

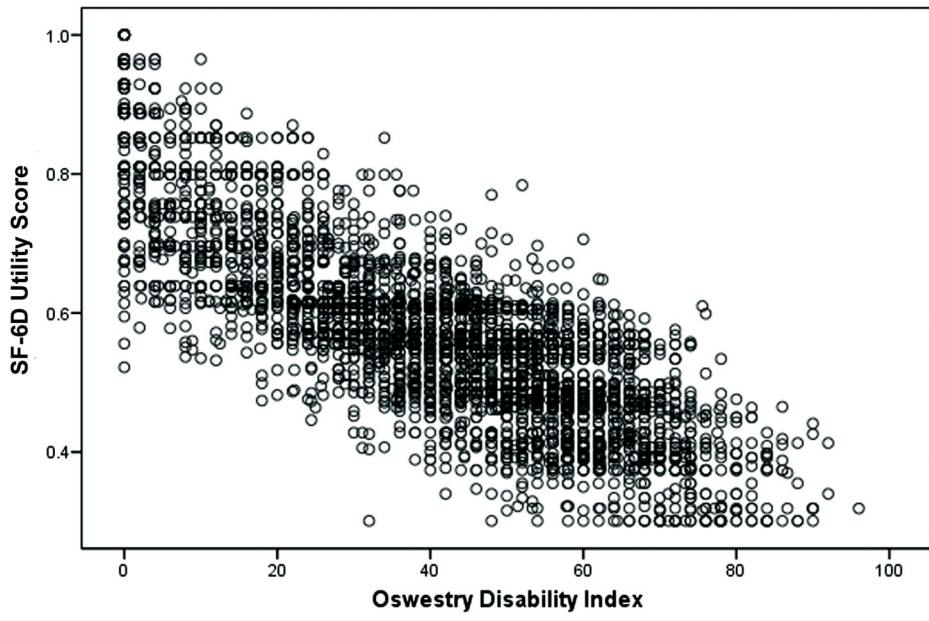


Figure 1.
Plot of ODI scores vs SF-6D scores

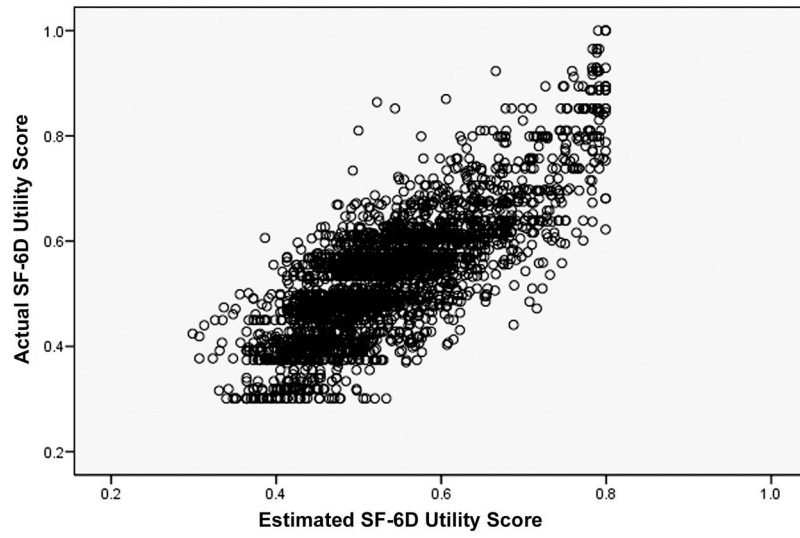


Figure 2. Estimated and Actual SF-6D for Low Back 2004, using model developed from Low Back 2000.

Table 1

Summary statistics of all paired observations from the LB2000 dataset

	SF-6D	ODI	BPNS	LPNS
Mean	0.57	40.54	5.98	5.54
SD	0.13	20.91	2.99	3.35
Minimum	0.30	0	0	0
Maximum	1	96	10	10

Table 2

Summary statistics all paired observations from the LB2004 dataset

	SF-6D	ODI	BPNS	LPNS
Mean	0.55	44.48	6.16	5.94
SD	0.12	19.75	2.83	3.04
Minimum	0.32	0	0	0
Maximum	1	90	10	10

Table 3

Mean scores for SF-6D, BPNRS and LPNRS across the different Oswestry Disability categories

Oswestry Disability Category	Mean SF-6D (SD)	Mean BPNRS (SD)	Mean LPNRS (SD)
Minimal Disability (0–20)	0.73 (0.11)	2.03 (1.81)	1.44 (1.96)
Moderate Disability (21–40)	0.60 (0.08)	4.73 (2.29)	4.00 (2.88)
Severe Disability (41–60)	0.52 (0.07)	6.79 (2.03)	5.97 (2.81)
Crippled (61–80)	0.43 (0.07)	8.08 (1.66)	7.37 (2.52)
Bed-bound (81–100)	0.37(0.07)	8.49 (2.15)	8.06 (2.68)

Table 4

Comparison of Various Models Predicting the SF-6D

Stepwise Regression Results	R	R ²	Adjusted R ²	RMSE
Oswestry Alone	0.82	0.67	0.67	0.07825
Oswestry plus Back Pain	0.83	0.69	0.69	0.07825
Oswestry, Back Pain, Leg Pain	0.83	0.69	0.69	0.07559