Conversion of PROMIS global health to EQ-5D health state values in patients undergoing lumbar spine surgery: A psychometric evaluation

Justin Turcotte
Maura Callanan
Brooke M. Buckley
Sohail Zaidi
Chad Patton

Follow this and additional works at: https://scholarlycommons.henryford.com/surgery_articles
Conversion of PROMIS global health to EQ-5D health state values in patients undergoing lumbar spine surgery: A psychometric evaluation

Justin Turcotte, Maura Callanan, Brooke Buckley, Sohail Zaidi, Chad Patton

**ARTICLE INFO**

Keywords: PROMIS, EQ-5D, QALY, Psychometric, Spine surgery

**ABSTRACT**

This study seeks to validate the conversion of PROMIS-GH scores to EQ-5D Health Utility Index (HUI) values. Patients undergoing lumbar spine surgery were prospectively surveyed using EQ-5D-3L and PROMIS-GH short form instruments. EQ-5D-HUI scores, and PROMIS scores converted to HUI were calculated. Neither instrument reported a significantly higher mean HUI score using the EQ-5D compared to PROMIS-GH. Strong positive correlation and agreement were observed. Conversion of the PROMIS-GH to the EQ-5D-HUI appears to be viable for evaluating the health status of patients undergoing lumbar spine surgery.

1. Introduction

As the expenditure on U.S. healthcare services continues to rise at a projected growth rate of 5.5% annually, the value of medical and surgical interventions require increased scrutiny. Symptomatic spinal stenosis in the presence of lumbar degenerative spondylolisthesis (LDS) is a relatively common condition affecting approximately 6–13% of the population. The surgical management of spine pathologies presents a significant cost to society, as demonstrated by the 2011 finding that spinal fusion accounted for over $12 billion in annual cost to U.S. hospitals, the highest of any surgical procedure. The clinical effectiveness of surgery for LDS is well established, as patients undergoing surgical treatment have showed substantially greater improvement in pain and function over periods of two, four and eight years than those treated nonsurgically in a large prospective randomized controlled trial; however, additional cost-effectiveness studies of this population are needed. In order to assess the cost effectiveness and cost-utility of treatment options across disease states with varying outcomes, standardized measures of health status are required. The most commonly used measure of health status in economic analysis is the quality adjusted life year (QALY). The quantification of health status on a zero to one-hundred point scale, herein described as a health utility index (HUI), can be measured using a variety of generic patient reported outcome measures (PROMs) to calculate QALYs.

The Euroqol Five Dimensions of Health measured at three levels (EQ-5D 3L) is a preference-based measure frequently used for calculation of HUI scores given its ability to produce a single measure of health status. Considered the gold standard legacy measure of health status, the EQ-5D has been validated in multiple patient populations across differing geographies and disease states. Alternatively, the Patient Report ed Outcome Measurement Information System (PROMIS) Global Health (GH) instrument is a standardized, generic patient-reported measure of global health as measured by physical and mental health.

Given its strong psychometric performance, ability to be integrated into electronic health record systems, and flexibility in application, the use of PROMIS instruments has rapidly increased. Despite this uptake, a limitation of the PROMIS GH instrument is its inability to produce a HUI value for use in the calculation of QALYs for cost effectiveness analysis. To fill this gap, a variety of methodologies for calculating HUI values from PROMIS GH scores have been proposed, although none have been validated specifically for patients undergoing surgery for LDS. This study seeks to validate the conversion of the PROMIS GH scores to HUI values using a previously published methodology. If successful, this would enhance the utility of the PROMIS GH instrument for economic analysis of surgery for LDS and could potentially reduce the requirement for using multiple generic health status PROMs in spine clinics, thus reducing patient burden and enhancing efficiency.
2. Methods

Seventy patients undergoing surgery for LDS by two fellowship trained orthopedic spine surgeons at a regional medical center from April 2017 to December 2018 were prospectively surveyed after receiving institutional review board approval. All patients were over eighteen years old and underwent primary surgery for LDS; revisions and patients with congenital or isthmic spondylolisthesis were excluded. After providing consent, patients were mailed a paper survey consisting of the EQ-5D 3L and the PROMIS Global Health v1.2 Short Form. All patients were mailed a first survey no earlier than nine months postoperatively and no later than fifteen months postoperatively. If no response was received within sixty days, a second survey was sent. Patient surveys were paired with demographics and information about their surgery and entered into a secure database. Health index values were calculated from the EQ-5D 3L using the US based time trade off weightings. PROMIS physical and mental health component scores were calculated and converted to EQ-5D 3L index values using the methodology described by Revicki et al. Descriptive statistics were run, and various techniques including paired and independent sample T-tests, Wilcoxon Signed Rank test, Pearson product moment correlation, Bland Altman analysis, Cronbach’s Alpha, and intraclass correlation were employed to assess instrument coverage, validity, and reliability. Statistical significance was assess at the $\alpha = 0.05$ level. All statistical analysis was performed using SPSS version 26 (IBM, Armonk, NY).

3. Results

3.1. Descriptive statistics

Thirty-three (47%) of patients responded to the survey with an average response time of 355 $\pm$ 77 days. The mean EQ-5D HUI for the population was 0.856 with a standard deviation of 0.131, compared to a PROMIS GH HUI of 0.732 with a standard deviation of 0.105. Median EQ-5D HUI was 0.827 compared to 0.768 for PROMIS GH HUI. Both measurements resulted in negative skew, with EQ-5D demonstrating a slight level ($-0.352$) and PROMIS GH demonstrating a moderate level ($-0.726$). Negative kurtosis was observed for both distributions (EQ-5D HUI $= 1.006$, PROMIS GH HUI $= -0.276$) suggesting a platykurtic distribution of scores. Normality of the distributions was assessed using Kolmogorov-Smirnov and Shapiro-Wilk tests, histograms and Q-Q plots. Neither the EQ-5D HUI (KS $= 0.273$, $p < .001$; SW $= 0.841$, $p < .001$) nor PROMIS GH HUI (KS $= 0.188$, $p = .004$; SW $= 0.932$, $p = .041$) were normally distributed, although Q-Q plot alignment demonstrated relatively close alignment to the normal distribution (Fig. 1). A summary of descriptive statistics is presented in Table 1 and a histogram of the distributions is presented in Fig. 2.
3.2. Instrument coverage

Coverage was defined as the continuum of a specific trait that can be assessed reliably by an instrument\textsuperscript{21} and assessed by evaluation of the floor and ceiling effects of the two methodologies evaluated. Neither the EQ-5D HUI nor PROMIS GH HUI demonstrated any floor effect in this sample. The EQ-5D HUI demonstrated significantly higher ceiling effects, with 14 of 33 (42.4%) patients reporting a perfect health state of 1.0; no patients achieved a health status of 1.0 using the PROMIS HUI conversion methodology, but 1 (3.0%) reached the maximum possible converted HUI value of 0.877. Graphical depiction of this ceiling can be observed in Fig. 2 histograms and Fig. 3 boxplots.

3.3. Validity assessment

Convergent construct validity was assessed using multiple methods.
Paired sample T-Tests indicated patients reported a significantly higher mean HUI score using the EQ-5D instrument than PROMIS GH (avg. = 0.865 vs. 0.732, t = 9.050, p < .001). This was confirmed by one sample T-Test of the mean difference between scores (avg. = 0.132 ± 0.084, t = 9.050, p < .001). Given deviation from the normal distribution and small sample size, median differences were also assessed using the related-samples Wilcoxon Signed rank test. A significant difference in median scores using the two methods was found (standardized test statistic = 4.851, p < .001). Pearson product moment correlation was performed to assess the linear relationship between HUI scores. The correlation coefficient (r) was .73 (p < .001) showing a strong positive, significant correlation between HUI scores using the two methodologies. Agreement between measures was then assessed using Bland Altman analysis. The difference in HUI score using the alternative methodologies was plotted against the mean of the two measures and assessed against a 95% confidence interval around the mean (Fig. 4). One of 33 data points fell outside of the 95% agreement interval, suggesting strong agreement between the instruments. Linear regression was then conducted to assess the relationship between mean HUI score of the two instruments and difference in calculated HUI scores. No significant relationship was found (β = .247, t = 1.922, p = .064), indicating no proportional bias between instruments.

3.4. Reliability assessment

Internal consistency of responses to each instrument was assessed using Cronbach’s alpha. The EQ-5D had a Cronbach’s alpha of 0.821, demonstrating good internal consistency. Cronbach’s alpha assessment comparing the physical and mental health components of the PROMIS GH instrument resulted in a score of 0.890, also suggesting good internal consistency. Reliability of the HUI scores between instruments was then assessed. Cronbach’s alpha was 0.857. Intraclass correlation was conducted to evaluate the level of agreement between the two instruments. An intraclass correlation coefficient of 0.857 (p < .001) revealed strong positive agreement between instruments.

4. Discussion

Our study demonstrates that conversion of the PROMIS GH to EQ-5D HUI is a valid and reliable method quantifying health status in patients.
undergoing surgery for LDS. Specifically, conversion of the PROMIS GH to EQ-5D resulted in reduced ceiling effects, while displaying adequate convergent construct validity, agreement, and internal consistency when compared to the EQ-5D HUI score. However, the significantly lower absolute HUI scores produced by the conversion formula must be recognized, and comparisons between studies using the alternative methodologies must consider the impact of the relatively lower health index values.

Both the EQ-5D and PROMIS instruments have been validated in low back pain and spine surgery patient populations. In a prospective study of 325 patients undergoing low back surgery, Söltänen et al. evaluated the psychometric properties of the EQ-5D compared to the legacy Oswestry Disability Index (ODI)—the most commonly cited disease specific measure of health status for low back conditions—and concluded the instrument displayed strong validity, reliability and responsiveness, and is suitable for use in cost utility analysis of this population. The two components of the PROMIS GH, physical function and mental health, have been independently validated in orthopedic and spine patients. In comparison with the ODI, the PROMIS physical function scale has been found to have better coverage, as demonstrated by reduced floor effects compared to the ODI. Validity of the PROMIS physical function scale in patients with spinal conditions was concluded based on strong correlation with the ODI in a study of 1607 patients with back and leg pain in a spine clinic. The mental health component of the PROMIS GH instrument demonstrates high reliability, as measured by an internal consistency coefficient of 0.86. The measure’s validity has been established through correlation with disease-specific PROMIS measurements; the mental health component of the GH instrument correlates strongly with the PROMIS depression scale ($r = -0.71$). Construct validity has been established across various disease states through significant positive correlation with other disease-specific and generic measures of mental health.

In a cross-sectional study of 14,962 orthopedic patients using the PROMIS anxiety, depression, physical function and pain interference scale, Beleckas et al. concluded that 20% of new orthopedic patients exhibit anxiety levels that may require intervention, with an even higher rate of 32.4% apparent in spine patients. The study further concluded that PROMIS physical function scores correlate more strongly with PROMIS anxiety scores than depression scores, suggesting the utility of assessing this specific dimension of mental health in orthopedic patients.

The methodology for converting PROMIS GH to EQ-5D HUI values used in our study was established by Revicki et al., in 2009. Using a sample of 20,400 cases from the PROMIS Wave 1 population, the authors employed a series of linear regression models to derive the final linear equation for conversion that is presented in the official PROMIS GH v1.2 scoring manual. Using various items from the PROMIS GH instrument, 65% of the variance in EQ-5D HUI scores was explained, compared to 57% of the variance when using the PROMIS GH summary domain scores. This result was comparable to other conversions to EQ-5D index scores; Lawrence and Fleishman explained 61–63% of EQ-5D index variance using the SF-12 summary score, while other methodologies using a variety of items explained 58–63% of the index values.

Differences between the predicted and actual index scores were between 0 and 0.02 points by gender and age groups, and an intraclass correlation coefficient of 0.77 between the actual and predicted scores indicated good agreement. Despite its inclusion in the PROMIS scoring guide, this conversion method has been critiqued for limitations including its underestimation of high HUI values and overestimation of low values as the equation only yields scores ranging from 0.328 to 0.877, thus resulting in artificial regulation to the mean.

Alternative methods of converting PROMIS scores to HUI values for economic analysis have been presented. In 2014, Craig et al. presented a method of converting PROMIS-29 responses directly into QALYs for cost-effectiveness research. Using a sample of 7557 respondents that was comparable to the 2010 US Census population distribution, the authors used multi-attribute utility regression and discrete choice experiments to derive the QALYs associated with various PROMIS-29 health states. In a 2017 follow up to this work, the authors presented an alternative model for mapping the PROMIS-29 to the EQ-5D 3L HUI scores based on responses from 2623 individuals completing both instruments. To account for the differences in scale, the authors transformed PROMIS-29 values to EQ-5D index scores using a transformation function derived from linear, power function, polynomial and piecewise linear regression models. After completing this conversion, the overall correlation between PROMIS 29 and EQ-5D 3L utility weights was 0.783. A final method for conversion using linear and equipercentile equating was presented by Thompson et al., in 2017. Using the PROMIS Wave 1 sample and the Cleveland Clinic neurological disease patient sample ($n = 6906$), the authors derived a series of models explaining 65–66% of the variance in actual EQ-5D index value, with differences of 0.01–0.02 between actual and predicted scores when stratified by age group. Notably, the authors applied the Revicki et al., 2009 equation to the dataset for direct comparison, and the updated equipercentile equation outperformed on multiple dimensions including overall predictive value, intraclass correlation coefficient and mean absolute error.

The authors did note the limitation of applying their model due to its increased mathematical complexity.

Our results deviate slightly from those establishing the utility of mapping the PROMIS GH instrument to EQ-5D HUI values. While these larger studies quote mean differences of approximately 0.01–0.02 between actual and calculated values, our sample found HUI values that were 0.132 ± 0.084 points lower than the actual EQ-5D HUI on average, a statistically significant difference. This result deviates significantly from a subset of 4854 spine patients in the Thompson study. Using the initial Revicki conversion equation, a difference of zero points was found between the actual (0.61 ± 0.22) and calculated (0.61 ± 0.10) HUI values.

Despite the significant difference in absolute scores observed, the PROMIS GH HUI did demonstrate positive association with the EQ-5D HUI across various measures. The correlation coefficient in our sample ($r = 0.73$, $p < .001$) was nearly identical to that of Revicki in the overall PROMIS Wave 1 population ($r = 0.785$). In addition, the ICC of 0.857 ($p < .001$), superior to that of 0.73 in the Revicki model. Further supporting the utility of the PROMIS GH HUI were the results of the Bland Altman analysis which indicated strong agreement between the instruments and no proportional bias. Finally, the strong reliability as demonstrated by Cronbach’s alpha assessment of PROMIS GH domains and with the overall scores provides further support for use of the methodology.

A potential benefit to using converted PROMIS GH HUI scores is the reduction of ceiling effects observed in EQ-5D HUI scores. While the Revicki methodology has been critiqued for its inability to produced high health state values, our results suggest the method has relatively low ceiling effects at 3% compared to the 42% of patients who reported a perfect health state using the EQ-5D. However, caution must be taken when comparing results using the alternative methodologies, as studies using the converted PROMIS GH converted HUI values will result in lower absolute levels of health status that will decrease the overall cost-effectiveness of interventions. Despite the limitations of the Revicki methodology, we find its simple and transparent formula attractive, and suggest it presents a low barrier for incorporation into standard clinical practice if a composite HUI is desired by practitioners. Further, its ability to use the PROMIS GH short form instrument rather than the 29-item PROMIS, its ease of use by alternative conversion formulas enhances its ability to be practically applied in time-constrained settings.

The primary limitation of our study is its small sample size. Given our sample population of only 33 patients, it is unsurprising that our results showed deviation from the normal distribution and previously published studies. Additionally, our study used paper forms of both the EQ-5D 3L and PROMIS Global Health v1.2 instruments. This has the potential to limit the applicability of our findings to patients using web based and alternative versions of these instruments such as the computer
adapted test. Despite these limitations, we feel the study holds significant value as it is the first evaluation of the conversion of PROMIS Global Health to EQ-SD health utility scores in this population to the authors’ knowledge.

5. Conclusion
Conversion of the PROMIS Global Health to EQ-SD health utility index value appears to be a viable option for evaluating the health status of patients undergoing surgery for spinal stenosis in the presence of lumbar degenerative spondylolisthesis. The convergent construct validity, reliability and coverage of the converted scores compare favorably to those of the actual EQ-SD HUI. Further study of the conversion methodology in a larger population of patients undergoing spine surgery is recommended before conversion of PROMIS Global Health to EQ-SD health utility index values is widely adopted for evaluation of health status in this population.

Funding
No funding was received for this study.

Ethical approval
All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent
Informed consent was obtained from all individual participants included in the study.

Declaration of competing interest
All authors declare they have no conflicts of interest.

References
30 Hartman JD, Craig BM. Comparing and transforming PROMIS utility values to the EQ-SD. Qual Life Res. 2017. https://doi.org/10.1007/s11136-017-1769-0.