Responsiveness of the EuroQoL 5-Dimension (EQ-5D) in Adolescent Idiopathic Scoliosis

ABSTRACT

Purpose: To test the responsiveness of the EuroQoL 5-dimension (EQ-5D) utility scores for adolescent idiopathic scoliosis (AIS).

Methods: A baseline sample of 227 AIS patients was recruited between August and October 2015, and was surveyed prospectively to 9-12 months follow-up. EQ-5D-5L utility scores were derived using a two-step approach: 1) cross-walking from 5-level responses to 3-level responses; and 2) applying the EQ-5D-3L Chinese population value set. An anchor approach was adopted to assess the responsiveness of EQ-5D. Effect size statistics (standardized effect size and standardized response mean) and independent t-test were used to assess the responsiveness, as well as to analyze the ability of measures to detect score changes with global health condition changes or discriminate between the worsened and unchanged/improved groups.

Results: Approximately two-thirds of follow-up patients (64.2%) reported no change in global health condition based on the self-reported health anchor, whilst 4.6% and 31.3% of patients rated worse and better in current health condition compared to baseline, respectively. In the subgroup where health worsened, EQ-5D utility scores were responsive to detect negative changes. EQ-5D utility scores had slight improvement in the group where health improved, despite a high mean score of 0.92 at baseline. Neither statistical

significance nor moderate-large effect size was observed in mean changes among unchanged group. Responsiveness property of the EQ-5D utility score was generally satisfactory with respect to each health condition group.

Conclusions: EQ-5D is found to be able to capture positive changes, and responsive in detecting important clinical changes in improved group of AIS population.

Key Words: Adolescent Idiopathic Scoliosis; Responsiveness; Anchor; EQ-5D; Utility

INTRODUCTION

Adolescent Idiopathic Scoliosis (AIS) is the commonest spine deformity which has no known etiology.[1,2] It is defined as a lateral curvature of the spine greater than 10 degrees accompanied by vertebral rotation, presenting between 10 and 18 years of age.[3] Curve progression can occur depending on a number of risk factors, such as the curve magnitude at initial presentation, which is suggested to be the most important predictor of long-term curve progression and behavior even beyond skeletal maturity.[4] Curve progression can lead to mental health concerns[5] with subsequent psychological issues and health-compromising behavior.[6] Interventions like bracing can span the entire pubertal period of these adolescents, and fusion surgery has a long-lasting effect that persists into adulthood.[7] Assessment of AIS patients' health-related quality of life (HRQoL) hence is an important factor that governs the outcome of interventions. In order to fully comprehend and interpret its long-term effect, HRQoL should not only be captured at a single time-point but also over time to gauge long-term outcomes.

One of the commonly used utility scores is the EuroQoL 5-dimension 5-level (EQ-5D-5L) questionnaire, which is a valid, reliable and sensitive generic measure to assess the HRQoL in AIS.[8] Despite having been previously validated and found to be an applicable outcome measure to be used in the local AIS population, the responsiveness of EQ-5D-5L has not been investigated. Responsiveness refers to the ability of a measure to detect clinically important changes in health over time[9], and hence crucial for calculating quality-adjusted life-years (QALYs) in cost-utility analyses.[10] Cost-utility assessment of treatment interventions in AIS is necessary to evaluate long-term outcomes.

Therefore, we hypothesize that the EQ-5D-5L mapped to the EQ-5D-3L utility scores is responsive to worsened and improved

health in AIS patients. The aim of this study is to evaluate the responsiveness of the EQ-5D instrument as a utility score for patients with AIS. This study is based on the global rating of change scale as an anchor, with calculation of effect size statistics among groups of patients with worsened, unchanged and improved health condition.

METHODS

This was a prospective study of AIS patients recruited from a scoliosis specialty clinic between August and October 2015. Ethics was approved by the local institutional review board. Consecutive sampling of AIS patients was performed at the initial visit and these subjects were surveyed at a follow-up of 9-12 months (between July and August 2016). Exclusion criteria included patients with non-idiopathic scoliosis (congenital/neuromuscular), who were illiterate and refused to participate. Patients who had given consent completed the EQ-5D-5L questionnaire (Hong Kong (traditional Chinese) EQ-5D-5L version)[11] at baseline in person at the clinic. The follow-up assessment was carried out in the form of a telephone interview conducted by single research personnel in random order. At follow-up interview, in addition to the EQ-5D-5L, the global rating of change scale was also completed. Subjects who were lost to follow-up were marked as defaulted. We planned for at least 50-100 samples as this was classified as good quality of evaluating responsive property as listed by the COSMIN checklist.[12]

Demographic data of the patients and clinical data at the time of initial visit were collected. All questionnaires were completed prior to the clinic consultation. The spine surgeon who provided such consultation was unaware of this study and assessed radiographs as usual. Whole spine standing radiographs (both posteroanterior and lateral) were taken on the same day as the consultation for

measuring the Cobb angle.[13] Scoliosis curves were classified by the modified Lenke classification[14] which included six curve types: type 1 (main thoracic), type 2 (double thoracic), type 3 (double major; thoracic curve larger than lumbar curve), type 4 (triple major), type 5 (thoracolumbar or lumbar curve), and type 6 (double major; thoracolumbar or lumbar curve larger than thoracic curve). Treatment modalities at initial visit of patients were retrieved from medical records and were recorded as: undergoing observation management, bracing, bracing followed by surgery and those who had previous fusion surgery undergoing regular review.

EuroQoL 5-dimension 5-level (EQ-5D-5L)

The EQ-5D-5L has five domain scales (mobility, self-care, usual activities, pain and discomfort, and anxiety and depression) and five levels for each domain. The EQ-5D-5L has 5 items, each digit in the five digit codes refers to the status of each dimension, ranging from 1 for no problem, to 5 for severe problem. Since the specific EQ-5D-5L value set / tariff for our country is currently not available, we adopted a two-step indirect approach to estimate EQ-5D-5L scores similar to another study.[15] The first step was the application of an indirect interim mapping method.[16] The EQ-5D-5L response values were transformed to the EQ-5D-3L response values according to the transition probability matrix. Subsequently, the EQ-5D-3L responsive values were scored according to a EQ-5D-3L value set ranging from −0.149 for the worst health status (‘33333’) to 1 for the full health (‘11111’).[17] The visual analogue scale (VAS) of the EQ-5D, in this case, had been omitted at the follow-up to avoid repetition and confusion, due to the introduction of the anchor item, which asked the patients to assess their own overall health condition over time.

Global Rating of Change Scale

For this study, an anchor, being an external criterion, was used as the reference to indicate patient improvement or worsening.[18] As for the AIS population whose cost-utility analysis is usually based on longitudinal follow-up, it is desirable to test the responsiveness of a transformation of the EQ-5D using an anchor of patient-reported assessment of health change over time (either prospectively or retrospectively determined) to indicate those for whom change in health occurred.[19]

The global rating of change (GRC) scale was therefore set as an anchor. GRC is a single-item outcome measure for independent scoring of self-perceived improvement in a patient, widely used for musculoskeletal research.[20] At the end of the follow-up assessment following the administration of EQ-5D-5L, GRC scale was administered to ask patients about their overall health condition as compared to baseline initial visit. The question posed was: “Compared to the first visit (9-12 months ago), how would you rate your overall health now?”. The response was a 7-point Likert scale: ‘extremely worse’ (rating of -3), ‘worse’ (rating of -2), ‘a little worse’ (rating of -1), ‘the same’ (rating of 0), ‘a little better’ (rating of +1), ‘better’ (rating of +2), ‘extremely better’ (rating of +3). Due to insufficient sample size in the three ‘worse’ subcategories, they were collapsed into a single ‘worsened’ subcategory. Similarly, collapse of three ‘better’ subcategories into a single ‘better’ subcategory was performed. The scale was then categorized into three meaningful health condition change groups: ‘worsened’ (rating from -3 to -1), ‘unchanged’ (rating of 0), and ‘improved’ (rating from +1 to +3).

Statistical Analysis

Differences in baseline characteristics between patients who have followed-up and defaulted were compared using independent t-test and Chi-squared test, where appropriate. These tests were carried out to test the indifference between the follow-up and defaulted subjects, so to eliminate any concerns of the sampling, response, and selection biases. Descriptive statistics including mean ± standard deviation (SD), ceiling and floor proportion of the EQ-5D-5L scores at baseline and follow-up assessments were reported.

The responsiveness of the EQ-5D instrument was assessed using the effect size statistics. The utility score difference between baseline and follow-up assessments were evaluated by standardized effect size (SES) and standardized response mean (SRM) separately for each group. The standard of SES and SRM was interpreted as trivial for values <0.2, small for values ≥0.2-<0.5, moderate for values ≥0.5-<0.8, or large for values ≥0.8, according to commonly accepted criteria.[21] The change in health condition was categorized into meaningful change groups for utility score comparisons.

In order to detect score changes with global health condition changes or discriminate between meaningful change groups, independent t-tests were performed to compare the utility score in patients with different groups of health condition changes. This enabled the assessment of the ability of the EQ-5D-5L instrument to match changes in utility score with health, and to discriminate among three groups of health conditions (worsened vs unchanged/improved; worsened/unchanged vs improved). All statistical analyses were conducted using STATA version 13.0. A p-value of <0.05 was considered as statistically significant. Multiple testing with Bonferroni correction was performed. 95% confidence intervals (CIs) were listed when appropriate.

RESULTS

The baseline characteristics of all patients are shown in **Table 1**. Out of a total of 227 patients recruited at baseline, 51 patients (22.5%) defaulted at follow-up assessment. At baseline, the majority of patients were female (74.9%), of mild or moderate curvature with Cobb angle of ≤40º (90.3%), and were under observation with regular follow-up (61.2%). For those who were prescribed bracing, 63% had already undergone bracing for at least one year or more. There were no significant differences in the characteristics of those who followed-up or defaulted except for the duration of bracing. Among the defaulted patients, 72.7% were patients with at least one year of bracing, who seemed more likely to default than patients with less than one-year bracing (27.3%). However, there were changes in the responses within each of the five domain scales of the EQ-5D-5L when comparing scores at the follow-up with the baseline as illustrated in **figures 1 and 2**.

Baseline and follow-up of EQ-5D-5L utility scores are shown in **Table 2**. Mean utility scores at baseline and follow-up assessments were 0.931±0.113 and 0.942±0.091, respectively, with insignificant mean change of 0.003±0.120 over time. There was no floor effect for the EQ-5D-5L utility score at baseline and follow-up but severe ceiling effects were observed. **Table 3** shows distribution of global rating of change scale. About two-third of follow-up patients (64.2%) reported no change in global health condition based on the self-reported health anchor, whilst 4.6% and 31.3% of patients rated worse and better in current health condition compared to baseline, respectively.

Table 4 illustrates the mean changes and effect size statistics for each health condition change group. In the subgroup where health worsened, EQ-5D-5L utility scores were responsive to detect negative changes. For those with health improvement, EQ-5D-5L utility scores also had slight improvements despite an already high mean score of 0.92 at baseline. Neither statistical significance nor

moderate-large effect size was observed with mean change in the unchanged group. **Table 5** shows the mean difference in change between health groups. Mean change over time in the improved group was statistically different from the worsened (P=0.033) and the worsened/unchanged (P=0.033) groups, whilst significant mean differences between improved/unchanged versus worsened group (P=0.038) was observed.

Responsiveness property of the EQ-5D-5L utility score was generally satisfactory with respect to all health condition groups.

DISCUSSION

AIS is the commonest spinal deformity that affects adolescents from puberty until skeletal maturity but the lasting effects of the disorder can reach adulthood. Hence, with possible interventions spanning a long period of time, decisions for observation (actively monitoring) versus bracing or surgical interventions may vary over time. There are reports of relative differences in properties among disease-specific measures such as SRS-7 versus SRS-22 in AIS children who had undergone spinal fusion, including their abilities in detecting changes in HRQoL measures preoperatively versus 1-year post-operation.[22] Similarly, reports of SRS-22[23] and SRS-30[24] changes 2-years after surgery and at 5-years[25] follow-up have also showed good reflection of HRQoL outcomes. However, it is still unknown how utility scores can assess the effect of scoliosis and its treatment outcomes on quality of life over time.

A recent study showed that the SRS-22 scores can be mapped onto EQ-5D-5L utility scores in AIS patients to generate cost-utility data.[26] However, this can only provide individual score sets without responsiveness information. This is the first study, to our best knowledge, that examines the responsiveness of EQ-5D in the AIS patients. We are in search of a generic instrument which can

generate utility scores, not only permitting comparison between patient groups,[27] but is essential in cost-utility analyses for longitudinal cohorts. It is necessary to analyze this tool's responsiveness and its ability to detect any improvement or worsening in quality of life. The responsiveness of EQ-5D has been studied in stroke, breast cancer and chronic obstructive pulmonary disease[28-30], however it has not been tested in an AIS population.

In this study, EQ-5D successfully captured positive and negative changes that matches improvement or worsening of health as suggested by the SES and SRM. Furthermore, the SES was near 0 in the unchanged group verifying its accuracy in detecting change. The EQ-5D has been found to be reasonably responsive in the worsened group, with SES of 0.71, comparable to that of 0.69 in breast cancer patients, with both studies adopting a self-rated change in quality of life.[29] Change in EQ-5D utility scores was able to differentiate between each of the health condition groups derived from the self-reported health anchor, except for comparison between unchanged and worsened group (P=0.216), as well as between improved and unchanged group (P=0.185). The mean changes in EQ-5D scores were statistically different at a significant level between improved versus worsened, improved versus worsened/unchanged, and improved/unchanged versus worsened groups. As generic instruments are designed to capture all aspects of HRQoL, they can provide a broader context by which to interpret the information about change in HRQoL[31] and QALYs in cost-utility analyses. Current results revealed that the ability of EQ-5D utility scores in detecting deterioration was better than detecting improvement, leading to the greater extent of QALYs loss in cost-utility analyses.

It may also be worthwhile to appreciate the changes within the five domain scales of EQ-5D-5L over time. These changes in various aspects can possibly contribute to the differences in the global health condition scale, given that the follow-up population was

comprised of very similar characteristics except the 2.5% increase in patients braced less than one year from baseline. The changes in EQ-5D-5L included the complete diminishing of 'moderate problems' in the Mobility, Self-care and Usual Activity domains while an increased proportion of patients shifted from 'moderate problem' to 'no problem' in the Anxiety domain, despite a higher proportion of both 'slight problem' and 'moderate problem' with Discomfort. The only aspect worsened was the discomfort at 9-12 months follow-up, whereas all other aspects had improved. How these 5 domains contributed to the worsened/improved/unchanged general health condition on the global health scale is of further interest, especially relating to different treatment modalities. The construct of the 5 domains enables EQ-5D-5L to be more receptive in detecting changes over time. In addition, EQ-5D-5L was able to still detect changes beyond the already very high ceiling effect at baseline, as reflected by the further increase in ceiling effect by 1.3% in the follow-up patients. High ceiling effect of EQ-5D-5L dimension might be in part explained by the generic nature of EQ-5D scale. This must take into consideration the 64.2% of patients who reported 'the same' in the global health condition, with a total of 4.6% reporting 'a little worse' or 'worse', and a sum of 31.3% with a response of 'a little better', 'better' or 'extremely better'.

The limitations of this study include the use of an indirect interim mapping method for EQ-5D-5L, instead of a direct valuation approach as the value sets for EQ-5D-5L are still under development.[32] There were contrasting claims however, that data sets generated by algorithm mapping method were found to be narrower than the time trade-off value sets,[33] whereas another study based on breast cancer data found a lack of differences including responsiveness of EQ-5D-5L scores by both approaches.[34] Another limitation was the different modes of conducting the EQ-5D-5L questionnaires at baseline and at follow-up, being filling up in person versus phone interview respectively. This was due to the timing of follow-up visits not coinciding with the responsiveness

time-point and it was impractical to recall patients only for the questionnaire. Possible discrepancies between written and verbal interviews were minimized by having the same research personnel conduct all phone interviews, in a systematic manner by which all 5 levels of responses of the EQ-5D-5L were read out to the subjects before an answer was given. Nevertheless, differential modes of administration have not been shown to lead to significant differences in EQ-5D scores.[35] Moreover, future study can also be improved with the use of multiple independent anchors (e.g. clinician-based anchor, proxy-based anchor versus patient-reported anchor) and to examine and confirm responsiveness across multiple samples. Despite the advantages of GRC scale being simple and widely used, weaknesses including less reliability and validity of single-item global rating of change scale were pointed out in previous literature.[20] Of note, this study made prior assumptions that changes in health condition resulted from scoliosis. Variations in EQ-5D-5L utility scores may not fully represent changes in health condition as a direct result of scoliosis. There may also be variations with different types of scoliosis and ethnic groups which require further study. Nonetheless, this study has shown that the EQ-5D is successful in detecting changes in health of an AIS population, and can serve as the basis of larger sample-sized, longitudinal study in the AIS population in the future to detect any significant changes over time for treatment modality in details, as well as to detect long-term outcomes.

CONCLUSION

The EQ-5D utility score is found to be able to capture positive changes in HRQoL score in AIS patients, and responsive in detecting important clinical changes in improved group of scoliosis population. Future studies about responsiveness property of EQ-

5D score with respect to health state deterioration are warranted.

REFERENCES

[1]. Miller NH (1999) Cause and natural history of adolescent idiopathic scoliosis. The Orthopedic clinics of North America 30:343-52, vii

[2]. Lonstein JE, Bradford DS, Winter RB, J O. Idiopathic scoliosis Moe's Textbook of Scoliosis and Other Spinal Deformities. 3rd ed: Philadelphia: WB Saunders; 1995.

[3]. Reamy BV, Slakey JB (2001) Adolescent idiopathic scoliosis: review and current concepts. American family physician 64:111-6

[4]. Tan KJ, Moe MM, Vaithinathan R, Wong HK (2009) Curve progression in idiopathic scoliosis: follow-up study to skeletal maturity. Spine 34:697-700

[5]. Schreiber S, Parent EC, Moez EK, Hedden DM, Hill D, Moreau MJ, Lou E, Watkins EM, Southon SC (2015) The effect of Schroth exercises added to the standard of care on the quality of life and muscle endurance in adolescents with idiopathic scoliosis-an assessor and statistician blinded randomized controlled trial: "SOSORT 2015 Award Winner". Scoliosis 10:24

[6]. Payne WK, 3rd, Ogilvie JW, Resnick MD, Kane RL, Transfeldt EE, Blum RW (1997) Does scoliosis have a psychological impact and does gender make a difference? Spine 22:1380-4

[7]. Weinstein SL, Ponseti IV (1983) Curve progression in idiopathic scoliosis. The Journal of bone and joint surgery American volume 65:447-55

[8]. Cheung PW, Wong CK, Samartzis D, Luk KD, Lam CL, Cheung KM, Cheung JP (2016) Psychometric validation of the

EuroQoL 5-Dimension 5-Level (EQ-5D-5L) in Chinese patients with adolescent idiopathic scoliosis. Scoliosis Spinal Disord 11:19

[9]. Guyatt G, Walter S, Norman G (1987) Measuring change over time: assessing the usefulness of evaluative instruments. Journal of chronic diseases 40:171-8

[10]. Whitehead SJ, Ali S (2010) Health outcomes in economic evaluation: the QALY and utilities. British medical bulletin 96:5-21

[11]. EuroQol EQ-5D-5L. https://euroqol.org/eq-5d-instruments/eq-5d-5l-available-modes-of-administration/self-complete-on-paper/. Accessed August 2015

[12]. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 21:651-7

[13]. Cobb J (1948) Outline for the study of scoliosis. In: Blount WP, Banks SW, eds. The American Academy of Orthopaedic Surgeons Instructional Course Lectures 5th Ed Ann Arbor, MI: JW Edwards 261-75

[14]. Sanders JO, Browne RH, McConnell SJ, Margraf SA, Cooney TE, Finegold DN (2007) Maturity assessment and curve progression in girls with idiopathic scoliosis. J Bone Joint Surg Am 89:64-73

[15]. Pan CW, Sun HP, Wang X, Ma Q, Xu Y, Luo N, Wang P (2015) The EQ-5D-5L index score is more discriminative than the EQ-5D-3L index score in diabetes patients. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 24:1767-74

[16]. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard AS (2012) Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in health : the journal of the International

Society for Pharmacoeconomics and Outcomes Research 15:708-15

[17]. Liu GG, Wu H, Li M, Gao C, Luo N (2014) Chinese time trade-off values for EQ-5D health states. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 17:597-604

[18]. Blum L, Korner-Bitensky N (2008) Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. Physical therapy 88:559-66

[19]. Cella D, Bullinger M, Scott C, Barofsky I (2002) Group vs individual approaches to understanding the clinical significance of differences or changes in quality of life. Mayo Clinic proceedings 77:384-92

[20]. Kamper SJ, Maher CG, Mackay G (2009) Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 17:163-70

[21]. J. C Cohen J. Statistical Power Analysis for Behavioral Science. 2. Hilsdale, NJ: Lawrence Earlbaum Associates; 1988. .

[22]. Jain A, Sponseller PD, Negrini S, Newton PO, Cahill PJ, Bastrom TP, Marks MC (2015) SRS-7: A Valid, Responsive, Linear, and Unidimensional Functional Outcome Measure for Operatively Treated Patients With AIS. Spine 40:650-5

[23]. Bennett JT, Samdani AF, Bastrom TP, Ames RJ, Miyanji F, Pahys JM, Marks MC, Lonner BS, Newton PO, Shufflebarger HL, Yaszay B, Flynn JM, Betz RR, Cahill PJ (2017) Factors affecting the outcome in appearance of AIS surgery in terms of the minimal clinically important difference. Eur Spine J 26:1782-8

[24]. Rodrigues LMR, Gotfryd AO, Machado AN, Defino M, Asano LYJ (2017) Adolescent idiopathic scoliosis: surgical treatment and quality of life. Acta Ortop Bras 25:85-9

[25]. Akazawa T, Kotani T, Sakuma T, Minami S, Torii Y, Orita S, Inage K, Fujimoto K, Shiga Y, Inoue G, Miyagi M, Saito W, Ohtori S, Niki H (2017) Midlife changes of health-related quality of life in adolescent idiopathic scoliosis patients who underwent spinal fusion during adolescence. Eur J Orthop Surg Traumatol

[26]. Wong CKH, Cheung PWH, Samartzis D, Luk KD, Cheung KMC, Lam CLK, Cheung JPY (2017) Mapping the SRS-22r questionnaire onto the EQ-5D-5L utility score in patients with adolescent idiopathic scoliosis. PLoS One 12:e0175847

[27]. Revicki DA, Leidy NK, Brennan-Diemer F, Sorensen S, Togias A (1998) Integrating patient preferences into health outcomes assessment: the multiattribute Asthma Symptom Utility Index. Chest 114:998-1007

[28]. Golicki D, Niewada M, Karlinska A, Buczek J, Kobayashi A, Janssen MF, Pickard AS (2015) Comparing responsiveness of the EQ-5D-5L, EQ-5D-3L and EQ VAS in stroke patients. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 24:1555-63

[29]. Lee CF, Luo N, Ng R, Wong NS, Yap YS, Lo SK, Chia WK, Yee A, Krishna L, Wong C, Goh C, Cheung YB (2013) Comparison of the measurement properties between a short and generic instrument, the 5-level EuroQoL Group's 5-dimension (EQ-5D-5L) questionnaire, and a longer and disease-specific instrument, the Functional Assessment of Cancer Therapy-Breast (FACT-B), in Asian breast cancer patients. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 22:1745-51

[30]. Nolan CM, Longworth L, Lord J, Canavan JL, Jones SE, Kon SS, Man WD (2016) The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. Thorax 71:493-500

[31]. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C (2003) Comparative responsiveness of generic and specific quality-of-life instruments. Journal of clinical epidemiology 56:52-60

[32]. Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro F (2014) A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 17:445-53

[33]. Golicki D, Niewada M, Hout Bv, Janssen MF, Pickard AS Interim EQ-5D-5L Value Set for Poland: First Crosswalk Value Set in Central and Eastern Europe. Value in Health Regional Issues 4:19-23

[34]. Luo N, Cheung YB, Ng R, Lee CF (2015) Mapping and direct valuation: do they give equivalent EQ-5D-5L index scores? Health and quality of life outcomes 13:166

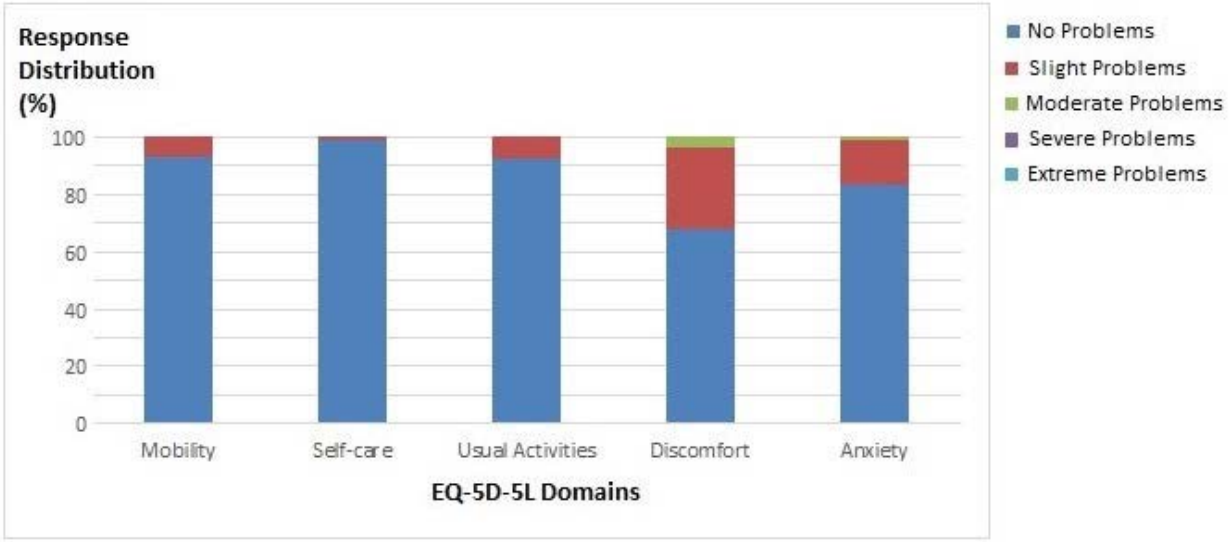
[35]. Chatterji R, Naylor JM, Harris IA, Armstrong E, Davidson E, Ekmejian R, Descallar J (2017) An equivalence study: Are patient-completed and telephone interview equivalent modes of administration for the EuroQol survey? Health Qual Life Outcomes 15:18

Figure Legends

Figure 1: The 5-level response distribution (%) of five domains in EQ-5D-5L at baseline.

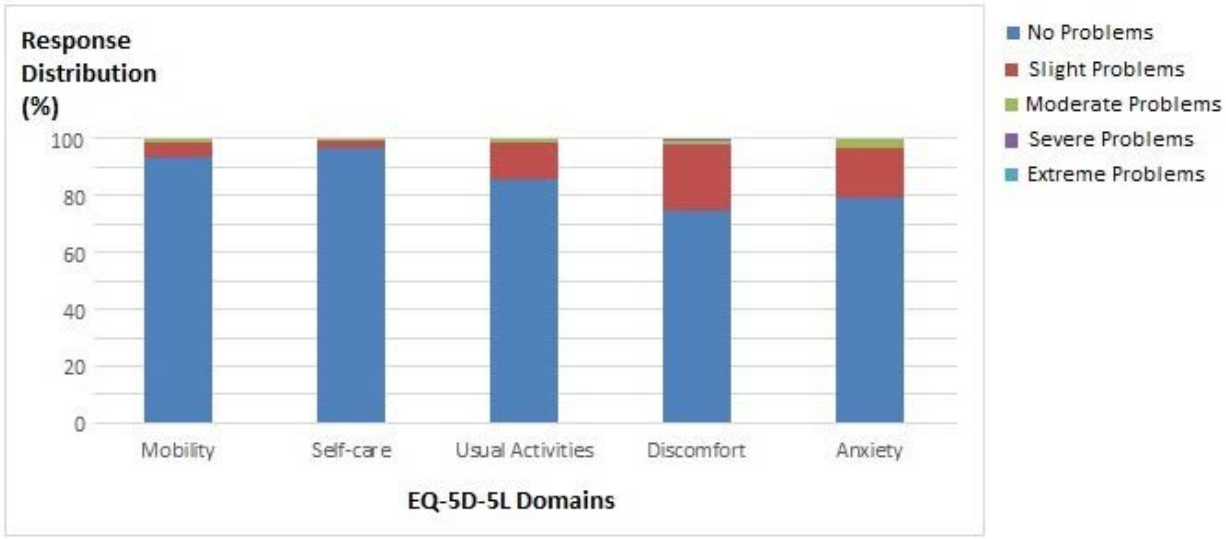
Figure 2: The 5-level response distribution (%) of five domains in EQ-5D-5L at 9 – 12 months follow-up. At follow-up, response 'Moderate problem' (in light green) completely diminished in the domains Mobility, Self-care and Usual Activities, whereas increase in responses 'Slight problem' and 'Moderate problem' in Discomfort was noted.

1 Figure 1



2

3 Figure 2


Response Distribution (%)
100
80
60
40
20
0
Mobility
Self-care
Usual Activities
Discomfort
Anxiety
EQ-5D-5L Domains
No Problems
Slight Problems
Moderate Problems
Severe Problems
Extreme Problems

1

2

Table 1. Baseline Characteristics of Adolescent Idiopathic Scoliosis Patients who complete and did not complete assessment at follow-up

Characteristics	Baseline (N=227)			Follow-up (N=176)			Default (N=51)	
	N	%	EQ-5D-5L score (Mean±SD)	N	%	EQ-5D-5L score (Mean±SD)	N	%
Age (years, Mean±SD)	15.6 ± 4.5			15.8 ± 4.9			14.8 ± 2.3	
<18	193	85.0 %	0.927 ± 0.118	145	82.4 %	0.933 ± 0.112	48	94.1 %
≥18	34	15.0 %	0.957 ± 0.071	31	17.6 %	0.961 ± 0.071	3	5.9 %
Gender								
Male	57	25.1 %	0.939 ± 0.104	45	25.6 %	0.954 ± 0.085	12	23.5 %
Female	170	74.9 %	0.929 ± 0.116	131	74.4 %	0.937 ± 0.093	39	76.5 %
Cobb Angle (Mean±SD)	25.0 ± 11.4			25.2 ± 11.4			24.6 ± 11.8	
≤40° (Mild or moderate)	205	90.3 %	0.934 ± 0.112	159	90.3 %	0.944 ± 0.091	46	90.2 %
>40° (Severe)	22	9.7 %	0.912 ± 0.123	17	9.7 %	0.917 ± 0.097	5	9.8 %
Treatment modality								
Observation with regular follow-up	139	61.2%	0.961 ± 0.069	104	59.1%	0.947 ± 0.085	35	68.6%
Braced before	13	5.7 %	0.974 ± 0.067	12	6.8 %	0.915 ± 0.128	1	2.0 %
Bracing	54	23.8 %	0.874 ± 0.146	43	24.4 %	0.926 ± 0.099	11	21.6 %
Surgery	21	9.3%	0.860 ± 0.175	17	9.7%	0.966 ± 0.078	4	7.8%
Duration of bracing								
<1 year	20	37.0 %	0.815 ± 0.145	17	39.5 %	0.935 ± 0.090	3	27.3 %
≥1 year	34	63.0 %	0.908 ± 0.137	26	60.5 %	0.921 ± 0.105	8	72.7 %
Modified Lenke Classification								
Type 1	63	27.8 %	0.918 ± 0.135	46	26.1 %	0.915 ± 0.089	17	33.3 %
Type 2	23	10.1 %	0.922 ± 0.102	19	10.8 %	0.919 ± 0.107	4	7.8 %
Type 3	41	18.1 %	0.941 ± 0.124	34	19.3 %	0.955 ± 0.086	7	13.7 %
Type 4	11	4.9 %	0.962 ± 0.085	5	2.8 %	0.949 ± 0.070	6	11.8 %
Type 5	38	16.7 %	0.925 ± 0.106	29	16.5 %	0.965 ± 0.085	9	17.7 %
Type 6	51	22.5 %	0.944 ± 0.088	43	24.4 %	0.953 ± 0.093	8	15.7 %

N = number of subjects; SD = Standard deviation; EQ-5D-5L = EuroQol Five Dimensions Five-level

Table 2. Descriptive statistics of EQ-5D-5L utility score at baseline and follow-up

	Mean	Standard deviation	Observed Range	Theoretical Range	Floor (%)	Ceiling (%)
EQ-5D-5L						
Baseline	0.931	0.113	0.339-1.000	-0.149-1.000	0%	65.6%
Baseline*	0.939	0.106	0.505-1.000	-0.149-1.000	0%	69.7%
Follow-up	0.942	0.091	0.610-1.000	-0.149-1.000	0%	66.9%
Mean Change	0.003	0.120				

EQ-5D-5L = EuroQol Five-Dimensions Five-level

Note:

*Baseline descriptive statistics of respondents who have completed both baseline and follow-up

Table 3. Distribution of global rating of change scale

Response	Follow-up (n=176)	
	N	%
-*3* extremely worse	0	0.0 %
-*2* worse	1	0.6 %
-*1* a little worse	7	4.0 %
0 same	113	64.2 %
1 a little better	20	11.4 %
2 better	18	10.2 %
3 extremely better	17	9.7 %

Table 4. Mean Change, Standardized Effect Size and Standardized Response Mean of EQ-5D-5L Utility Scores by Global Rating of Change Scale

	Mean (±SD) at baseline	Mean (±SD) at 6-month follow-up	Mean Change (±SD)†	p-value	SES (95% CI)	SRM (95% CI)
Worsened group (n = 8)	0.92±0.12	0.84±0.14	-0.08±0.21	0.299	-0.71 (-1.81,0.43)	-0.40 (-1.02,0.24)
Unchanged group (n = 113)	0.95±0.09	0.95±0.09	-0.01±0.10	0.579	-0.06 (-0.25,0.15)	-0.05 (-0.23,0.13)
Improved group (n=54)	0.92±0.13	0.95±0.09	0.03±0.13	0.093	0.25 (-0.01,0.57)	0.23 (-0.01,0.54)

SES=standardized effect size; SRM=standardized response mean; CI = Confidence Interval; SD=standard deviation

Note:

† Higher scores represents a higher level of functioning or a better HRQOL

Table 5. Mean difference in change of EQ-5D-5L utility scores with 95% confidence interval on discriminating groups of global health condition

	Mean difference in change (95% CI)	p-value
Unchanged - worsened group†	0.08 (-0.026 - 0.182)	0.216
Improved - worsened group†	0.11* (0.007 - 0.222)	0.033
Improved - unchanged group†	0.04 (-0.010 - 0.084)	0.185
Improved - worsened/unchanged group	0.04* (0.004 - 0.080)	0.033
Improved/unchanged – worsened group	0.09* (0.005 - 0.174)	0.038

CI = Confidence Interval

Note:

*Significant difference between groups with p-value < 0.05

†Tested by Bonferroni correction

Table 6. Factors associated with change in EQ-5D-5L utility scores by multiple linear regressions

Factor†	Change in EQ-5D-5L score (n = 175)		
	Coeff.	95%CI	p-value
Age, year	-0.004	(-0.008,0.000)	0.072
Female (vs. Male)	-0.028	(-0.068,0.012)	0.165
Cobb Angle	0.001	(-0.001,0.002)	0.434
Treatment modality (vs. Observation with regular follow-up)			
Braced before	-0.030	(-0.104,0.044)	0.424
Bracing	0.062*	(0.020,0.105)	0.004*
Surgery	0.103*	(0.041,0.164)	0.001*
Modified Lenke Classification (vs. Type 1)			
Type 2	-0.008	(-0.071,0.054)	0.793
Type 3	-0.005	(-0.057,0.047)	0.851
Type 4	0.036	(-0.072,0.143)	0.513
Type 5	0.042	(-0.012,0.097)	0.127
Type 6	0.012	(-0.037,0.061)	0.623

EQ-5D-5L = EuroQol Five-Dimensions Five-level; CI = Confidence Interval; Coeff = Coefficient

Notes:

*Significant with p-value < 0.05

†Variable in brackets is the reference category for independent variables