SPINE An International Journal for the study of the spine, Publish Ahead of Print

DOI: 10.1097/BRS.0000000000003431

Feasibility of Proxy-reported EQ-5D-3L-Y and Its Agreement in self-reported EQ-5D-3L-Y for Patients with Adolescent Idiopathic Scoliosis

Jiaer Lin[1], Carlos King Ho Wong, PhD[2], Prudence Wing Hang Cheung, BDSc (Hons)[3], Nan Luo, PhD[4], Jason Pui Yin Cheung, MBBS, MMedSc, MS, PDipMDPath[3]

[1] Jockey Club School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong SAR, China

[2] Department of Family Medicine and Primary Care, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

[3] Department of Orthopaedics and Traumatology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

[4] Saw Swee Hock School of Public Health, National University of Singapore, Singapore

Corresponding Authors:

Dr. Carlos KH Wong
Department of Family Medicine and Primary Care
The University of Hong Kong
Rm 1-01, 1/F, Jockey Club Building for Interdisciplinary Research, 5 Sassoon Road, Pokfulam, Hong Kong SAR, China
Tel.: (852) 28315055, Fax No.: (852) 28147475, Email: carlosho@hku.hk

Dr. Jason Pui Yin Cheung
Department of Orthopaedics and Traumatology
The University of Hong Kong
5/F Professorial Block, Queen Mary Hospital, Pokfulam, Hong Kong SAR, China
Tel: (852) 22554581, Fax: (852) 28174392, Email: cheungjp@hku.hk

The manuscript submitted does not contain information about medical device(s)/drug(s).

The EuroQol Research Foundation (grant number: 20180530) and the General Research Fund of the Research Grants Council (#17119518 and #17156416) funds were received in support of this work.
No relevant financial activities outside the submitted work.

Abstract

Study Design: Prospective cohort study

Objective: To compare feasibility of self-reported and proxy-reported EQ-5D-3L-Y, to estimate the agreement of health outcome between patients with AIS and their proxies, and to examine factors that may affect patient-proxy agreement.

Summary of Background Data: The EQ-5D-3L-Y questionnaire has both self-reported and proxy-reported versions. Despite previous studies have indicated that proxies tended to respond higher or lower levels of severity in specific dimensions than patients report, the level of agreement between children with AIS and their proxies remained unknown.

Methods: A consecutive sample of patients with AIS and their caregivers were recruited. Feasibility was tested according to the proportion of missing responses. Agreements between self-report and proxy EQ-5D-3L-Y were evaluated using perfect agreement (PA), Gwet's Agreement coefficients (AC) and the intra-class correlation coefficients (ICC). Linear regressions and logistic regressions were conducted to assess the factors associated with the agreement in health outcome between self-reported and proxy-reported EQ-5D-3L-Y.

Results: A total of 130 patient-proxy pairs was involved in the study. Agreement of EQ-5D-3L-Y responses between the self-report and proxy version was good for "Feeling worried/sad/unhappy" dimension, and very good for other dimensions. Poor agreement in VAS score was observed between patient and proxy versions. Proxy's education level, patient's MLC and treatment modality were the significant determinants of the agreement in in "Mobility", "usual activities" and "pain/discomfort" dimension respectively.

Conclusions: Proxy-reported EQ-5D-3L-Y demonstrates good feasibility and satisfactory agreement with patient version. Proxy's education appears to have positive influence in agreement between patient-proxy dyads.

Keywords: adolescent idiopathic scoliosis, feasibility, agreement, proxy, EQ-5D-3L-Y

Level of Evidence: 2

Key Points

• This study is the first to evaluate the feasibility of proxy-reported EQ-5D-3L-Y and to compare its agreement with self-reported version in adolescent idiopathic scoliosis populations worldwide.

• Proxy-reported EQ-5D-3L-Y demonstrates good feasibility and satisfactory agreement with patient version and proxy's education appears to have positive influence in agreement between patient-proxy dyads.

• This study results serve as basis and reference to choose self-reported or proxy-reported EQ-5D-Y measurement for clinicians and researchers, allowing for selecting appropriate interventions for patients with adolescent idiopathic scoliosis.

Introduction

Scoliosis, a lateral curvature of the spine greater than 10 degree as measured by Cobb method on a standing radiograph, is a 3-dimensional spinal deformity [1]. Adolescent idiopathic scoliosis (AIS) is the most common type of scoliosis which affects children aged 10 to 18 [2]. Patients present with marked deformity, psychosocial pressure and decreased self-confidence occurring with curve progression [3], impacting health-related quality of life (HRQoL) substantially [4]. Several treatments including observation, bracing and surgery also influence on HRQoL [5], defined as a multidimensional construct [6]. There is a need to apply HRQoL measurements to evaluate the AIS patients' health outcome to select more appropriate interventions.

WHO recommended that HRQoL measurements should at least involve physical, psychological and social health dimensions [7]. Both the refined Scoliosis Research Society-22 (SRS-22r) and EuroQoL Five-Dimension Three-Level Questionnaire (EQ-5D-3L) are commonly-used instruments for AIS patients [8,9]. SRS-22 is a disease-specific measurement [10] for evaluating the effects of interventions [11,12] whereas EQ-5D-3L is more generic, allowing the calculation of quality-adjusted life years (QALYs) in health economic evaluation. In 2009, the youth version of the EQ-5D-3L (EQ-5D-3L-Y) was developed [13] and its psychometric properties of the EQ-5D-3L-Y were found to be satisfactory in general [13] and disease-specific [14] child and adolescent populations [15].

The EQ-5D-3L-Y questionnaire has both self-reported and proxy-reported versions [16]. Although self-reported EQ-5D-Y is the most widely-used measure [8], proxy responses are needed when children are too young or immature to interpret the questions [17]. On behalf of patients, proxies can reduce missing data and improve the generalizability of study results to avoid decreasing participants [18,19]. Proxy HRQoL measurement showed overwhelming agreement with the patient version in Epilepsy populations [20] but proxies tended to respond higher or lower levels of severity in specific dimensions than patients report [21]. In a survey for the Japanese version of the EQ-5D-Y, proxies showed perfect agreement with patients in "Looking after myself" dimension but lowest in "Feeling worried or sad or unhappy" [22]. Thus it is important to ascertain the level of agreement between patients and proxies. Given that parents or employees reported better levels of agreement in specific dimensions [23 24], we inferred that patient and proxy factors may have an impact in patient-proxy agreement.

Therefore, the aims were to compare the feasibility of self-reported and proxy-reported EQ-5D-3L-Y to estimate the agreement of health outcome between patients with AIS and their proxies. Besides, this study examines factors that may affect patient-proxy agreement.

Methods

Participants

This was a prospective study of 130 patients who were consecutively recruited from the paediatric spine tertiary referral centre from December 2018 to January 2019. Caregivers who were most familiar with patients' health condition and behavior were invited to participate. The inclusion criteria for the patients was: 1) diagnosis of juvenile (JIS) or AIS, 2) children and adolescents aged between 10 to 18 years old. Before the clinic consultation, patients and their proxies were requested to complete the EQ-5D-3L-Y questionnaires at the same time, as well as the EQ-VAS and questions regarding sociodemographic factors including patients' age, gender, body mass index. Cobb angle, scoliosis curvature types and treatment modality were retrieved from medical records. Besides, proxies also reported their age and relationship with the patients and education level. Question that whether proxy lived with patients or not was also included. Ethical approval was obtained from the local ethics committee. Informed consent was obtained prior to data collection.

Study instrument

EQ-5D-3L-Y is an adaptation version of the EQ-5D-3L for children from 8 to 15 years old [25]. Pilot testing showed great responsiveness compared to the standard EQ-5D-3L [26]. The EQ-5D-3L-Y consists of a descriptive system with five dimensions and a visual analogue scale (VAS) [24]. Two proxy versions include "the proxy rates how he or she rates the health of the child" and "the proxy rates how he or she thinks the child would rate his or her own state if he or she were able to do so" [16]. This analysis adopted the former as the perspective of proxy-reported EQ-5D-3L-Y, the content of questionnaires for both patients and proxies was identical.

Statistical Analysis

Descriptive statistics were reported and independent t-test and Chi-squared test were performed where appropriate. Feasibility was determined by calculating the percentages of missing values which were defined as a sample which completely omitted a dimension of the questionnaire. Proportion of the "no problem" responses was calculated to assess the ceiling effects between patient and proxy and compared by McNemar test. Relative reduction ((Ceiling effects $_{\text{self-reported}}$ − Ceiling effects $_{\text{proxy-reported}}$)/Ceiling effects $_{\text{self-reported}}$) and absolute reduction from patient to proxy were also reported.

Agreement between self-reported and proxy-reported VAS scores was measured by the intra-class correlation coefficients (ICC) [27], with an ICC of <0.4 indicating poor agreement; 0.4-0.74 indicating moderate agreement; ≥0.75 indicating good agreement [27]. We calculated Gwet's Agreement coefficient (Gwet's AC) [28] and percent agreement (PA) to assess the inter-rater reliability, with poor agreement when Gwet's AC<0.2; moderate (0.21-0.4); good (0.61-0.8) and very good (over 0.8) [28]. Agreement evaluation was also compared by Chi-squared test on known subgroups: (1) Patient age: 10-12 years old versus 13-16 years old; (2) Patients Cobb angle: ≤40° versus >40°; (3) Patients treatment modality: observation versus bracing; (4) Patients curvature type: Thoracic curve versus Lumbar curve versus Thoracic & Lumbar curve; (5) Proxy relationship: mother versus non-mother; (6) Proxy age: <40 years old versus ≥40 years old; (7) Proxy education: secondary or below versus tertiary or university.

To further examine the magnitude of the association, we fitted separate linear regression models for self-reported, proxy-reported score and their score differences, logistic regression for patient-proxy dyad responses in each dimension. Since value sets for the EQ-5D-3L-Y are not yet available and adult value sets were proved to be potentially misleading

for children [8], the index score was not calculated. Instead we used EQ-VAS scale which had been validated to be consistent with the descriptive system of EQ-5D-3L-Y in general population [29]. As dependent variable, absolute difference means that unit increase in an independent variable is related to a reduced agreement when coefficient of the model is positive.

A P-value less than 0.05 was considered statistically significant and data analyses were conducted using SPSS Windows 24.0 (IBM SPSS Inc., Chicago, IL, USA) and the STATA software (StataCorp LP, College Station, Tex) version 13.0.

Results

Feasibility

130 targeted patients met the selection criteria and were approached to participate, together with their proxies. Of these, five patients and their proxies failed to fill out the EQ-5D-3L-Y and were excluded from the analysis. One patient failed to complete the self-reported questionnaire, and one missing response was found on the dimension of "Depression" reported by patients. There were less than 3% missing data in VAS score, the agreement on the VAS score was analyzed for 125 patient-proxy pairs. Descriptive statistics are presented in table 1. 90.6% patients were female, 86.1% had Cobb angle ≤40° and the median age was 14 (range 11-16) years old. 27% received observation management and 35% were prescribed a custom molded underarm thoraco-lumbo-sacral orthosis. Patients were suggested to use the brace full-time (>20 hours a day) but were not restricted for movement. Scoliosis specific exercises were suggested for all patients with regular follow-up with physiotherapists. Proxies were mostly mothers (70.0%), with the median age of 45 (range 32-90) years. Over 95% of proxies had received education and less than 4% indicated not living with patients.

Responses distribution

Table 2 shows the distribution of responses describing problems. The majority of patients and proxies reported no problems regardless of any dimensions. No patients rated the worst problem level on dimensions (33333), only one proxy reported patients' health as the worst possible except "Feeling worried, sad or unhappy" dimension and three reported extreme problems in this dimension. Figure 1 demonstrates that there was a tendency that "Some problems" was used more frequently by proxies than patients except for "Doing usual activities" dimension. There were significant differences in response distribution of "Mobility" with p-value <0.05. Both patient (99.2%) and proxy (97.7%) responses showed significant ceiling effects in the "Looking after myself" dimension, which was consistent with the results of earlier studies [30]. The comparison of ceiling effects was displayed in table 3. Overall, the proportion of patients perceived as "no problems" levels was higher than that of proxies except for "Doing usual activities", with significant differences in "Mobility".

Overall patient and proxy agreement

The level of agreement in dimension responses and EQ-VAS score between patient and proxy questionnaires is depicted in table 4. In general, agreement based on AC and PA was found to be higher in "Looking after myself", "Doing usual activities" and "Mobility" than in "Having pain or discomfort" and "Feeling worried, sad or unhappy". The study demonstrated good agreement in the "Feeling worried, sad or unhappy" dimension whereas all other dimensions had very good level of agreement.

Effects of factors on agreement

As shown in table 4, the agreement for subgroups was similar with overall participants across five dimensions except for the "Mobility" and "Having pain or

discomfort". Proxy's education level was observed to have statistical correlation ($p<0.05$) with agreement in the "Mobility" dimension while treatment modality significantly influenced agreement in the "Having pain or discomfort" dimension. Treatment modality was also statistically related with "Doing usual activities", so as curvature types, but the agreements were in the same level with the overall population (very good). It should be noted that agreement was detected to be perfect in "Looking after myself" when patients were aged 13 to 16 years or with observation treatment, thoracic curve or thoracic and lumbar curves, and when proxies were younger than 40 years. Besides, patients with severe curvature degrees had perfect agreement both in "Looking after myself" and "Doing usual activities". By far, the EQ-5D-3L-Y showed satisfactory agreement between patient and proxy version. However, the ICC for the EQ-VAS score was low regardless of overall or known groups.

Table 5 shows the results of regression models fitted to assess the association between participants' factors with agreement on EQ-VAS score and dimension responses. Patients living with proxies had most positive association (coefficient=6.1, 95%CI=-3.7,16.0) with absolute difference of EQ-VAS score between patient and proxy-reported versions. But no factors yielded significant correlations with agreement in EQ-VAS scores. "Looking after myself" dimension had perfect agreement in logistic regression. As compared to those with lower education level, proxies who received tertiary or university education had significantly larger odds (OR=9.4, 95%CI:1.0-87.2, $p<0.05$) of agreement in responses of "Mobility". The curvature type and treatment modality showed significant relationship with agreement. More specifically, significant negative association with agreement were observed in the thoracic and lumbar curve group (OR=0.1, 95%CI:0.0-0.7, $P<0.05$) and bracing treatment group (OR=0.3, 95%CI:0.1-0.9, $P<0.05$) in the "Doing usual activities" and "Having pain/discomfort" respectively.

Discussion

Previous studies have investigated the parent-child agreement in the general population [29] but none of studies explored the proxy version of EQ-5D-3L-Y in patients with AIS. This is the first prospective study to our knowledge evaluating the feasibility of proxy-reported EQ-5D-3L-Y and the agreement of health outcome between patient and proxy version in AIS populations. The following were the main findings of the research.

This study shows that the proxy version may slightly improve the feasibility of EQ-5D-3L-Y completed by patients with no missing data as observed in the individual dimensions. In fact, the EQ-5D-3L-Y has been proved to have excellent feasibility in the general population [13] and in the AIS population [30]. This finding suggests that both patient and proxy versions are highly feasible for evaluating the HRQoL in AIS patients. Both versions show high ceiling effect and no floor effect across all dimensions which may be partly explained by the generic nature of EQ-5D-3L-Y. Nevertheless, "Mobility" was the only dimension with significantly reduced proportion of "no problems" responses from self-report to proxy-report version. This finding provides the evidence for improving the application of proxy-reported EQ-5D-3L-Y as an essential supplement to the self-reported EQ-5D-3L-Y.

With regard to agreement, patients and proxies tended to have better agreement in observable behaviors ("mobility", "self-care" and "usual activities") than in "pain/discomfort" and "worried/sad" dimension [18,31,32]. This phenomenon may suggest that patients' physical problems are relatively easier to be assessed than emotional problems by proxy. This could also be seen when patients had more severe spinal deformity, indicating that proxy version can be well utilized especially when patients are in severe health status. Nevertheless, since both versions of EQ-5D-3L-Y facilitated good or above agreement despite significant difference of responses, proxy report can also be applied no matter the

severity of illness or treatment method used. However, due to the negative impact of severe curvature magnitude and bracing treatment on the agreement in "Pain or discomfort" and "Worried, sad or unhappy", researchers using the proxy version should note the discrepancies of responses in unobservable dimensions.

Of note, all subgroups demonstrated very good agreement in "Mobility" except for proxies who had lower education, reflecting that the lower-education-level proxies may prefer reporting problems, thus contributing to significant differences in ceiling effect between proxy-patient. Moreover, we found perfect agreement in patients aged 13 to 16 or proxies aged under 40 in self-care dimensions. These findings may be important information for re-defining the age criteria for proxy EQ-5D-3L-Y administration. Our study shows poor agreement in VAS score which is in line with other proxy assessment in stroke populations [33] due to less reliable properties of VAS scoring in children for general use [34].

Agreement in VAS responses between patient and proxy pairs was not driven by patient's age and gender [35-37] and patient's other sociodemographic and clinical factors, which was consistent with studies in the Cerebral palsy population[38] and a Japanese survey[22]. Nevertheless, noted that proxies who lived together with patients may under-evaluate patients' health by reporting lower EQ-VAS score than proxies who do not live with patients. A possible explanation is that proxies living with patients have more frequent contact with them. Hence, they were more sensitive to the routine behavioral impact of patients' illness like gait, feet or step position during walking, and unbalanced shoulder height. These are changes that may not be noticed by the patients themselves whereas proxies perceived it as an issue of patients' general health, and had more empathy for them as direct care-providers than those who do not live with the patient, contributing to lower health evaluation scores. In addition, girls reported higher scores than boys which was consistent with the previous study [21]. EQ-VAS score differences were bigger in girls. On the one hand, proxies may perceive

girls to be more vulnerable thus reporting lower scores than the boys' proxies. Because apart from the behavioral impact of the AIS, the disease has a significant effect on a patient's appearance, which may be perceived as more important for girls and thus proxies may have more concerns. On the other hand, female patients may be more able to endure pain than males[39], which may lead to higher self-reported EQ-VAS scores. We also found that proxies receiving tertiary or university education are in significantly better agreement with patients than those with lower education level in "Mobility" and has lower difference in patient-proxy VAS responses. As expected, this finding confirms the effect of education level on the "Mobility" dimension which is consistent with the ceiling effect and Chi-squared test results presented earlier.

Our study used Chi-squared analysis, multivariate linear regression and logistic regression to explore patient-proxy dyads agreement on responses by various subgroup factors. For dimension responses, both Chi-square test and logistic regression were used to identify factors that influence the agreement between child and proxy ratings of "no problem" (level-1), "some problems" (level-2) and "lots of problems"(level-3) for five dimensions. Chi-square test was a way to reveal the existent relationship between potential factors and the agreement based on approximations by comparing the expected counts with the observed counts of each problem level through projecting. Whereas logistic regression mainly aimed at testing the magnitudes of the correlations by controlling the confounders in the model. The dependent variable of logistic regression was the agreement status (1 refer to agreed and 0 refer to disagreed) after matching problem levels between patients and proxies, with level 1-1, 2-2, 3-3 answered by both patients and proxies considering as "agreed", other pairs were considered as disagreed. The significant factors of subgroups showed the same in the subgroup analysis through these two different methods, confirming the robustness and persuasiveness of the findings. Besides, multivariate linear regression was used to assess the

relationship between potential factors and EQ-VAS score. Because EQ-VAS is not preference-based, we were unable to justify if the proxy or patient over or under-evaluate patients' health outcome without knowing whose responses were truly reflecting the real score of patients' health conditions even though strong correlation was observed.

This study has several limitations. First, our study sample was mostly comprised of healthy children, which may affect the generalizability of our study results. It should be noted that although the proportion of patients with Cobb angle >40 degrees is only 7.5%, this is consistent with the small prevalence of patients with severe curves in a general AIS population[40]. Observation is the predominant treatment method for patients with mild curves according to the Scoliosis Research Society recommendations[2]. Hence, up to 71.3% of the patients were only observed without active treatment. Therefore, this is the represented population of patients with AIS. Second, our study patients aged between 10-18 years old according to AIS definition [2] whilst EQ-5D-3L-Y is designed for patients in the given age ranging from 8 to 15. However, the effect of age factor has been adjusted to the minimal because merely 12.6% of our patients are over the upper limit of 15 years old. Third, lack of child-specific value set for EQ-5D-3L-Y makes it not possible to compare the patient-proxy agreement in the utility measure which is essential for calculating QALYs in health economics evaluation. Fourth, results of the study cannot be generalized to the agreement over time in the absence of test-retest data. HRQoL of AIS patients has great changes after treatment like surgery which may normally take at least three months to recover, hence it is critical to investigate the agreement in different timepoints and in the magnitude of change in HRQoL.

Conclusion

In conclusion, proxy-reported EQ-5D-3L-Y demonstrates good feasibility and satisfactory agreement with the patient version. The proxy's education appears to have positive influence whereas patients with bracing of thoracic & lumbar curve have negative impact in agreement. This study results can serve clinicians and researchers of AIS as a basis and reference to choose HRQoL measurements, which in turn influence the selection for appropriate interventions.

Reference

1. Kane W. Scoliosis prevalence: a call for a statement of terms. *Clin. Orthop. Relat. Res.* 1977:43-6.
2. L Weinstein S, Dolan L, C Y Cheng J, et al. Adolescent idiopathic scoliosis. *Lancet* 2008;371:1527-37.
3. Asher M, Burton D. Adolescent idiopathic scoliosis: natural history and long term treatment effects. *Scoliosis* 2006;1:2.
4. Cheung PWH, Wong CKH, Cheung JPY. An Insight Into the Health-Related Quality of Life of Adolescent Idiopathic Scoliosis Patients Who Are Braced, Observed, and Previously Braced. *Spine* 2019;44:E596-E605.
5. Deceuninck J, Bernard J. Quality of life and idiopathic scoliosis treated by brace. Transversal study. *Scoliosis* 2010;5:O38.
6. D L P. Generic and disease-specific measures in assessing health status and quality of life. *Medical care* 1989;3 Suppl.
7. Constitution of the World Health Organization. 1946. *Bulletin of the World Health Organization* 2002;12.
8. Thorrington D, Eames K. Measuring Health Utilities in Children and Adolescents: A Systematic Review of the Literature. *PloS one* 2015;10:e0135672.
9. Oppenlander ME, Maulucci CM, Ghobrial GM, et al. Research in spinal surgery: Evaluation and practice of evidence-based medicine. *World journal of orthopedics* 2014;5:89.

10. Danielsson AJ, Romberg K. Reliability and validity of the Swedish version of the Scoliosis Research Society-22 (SRS-22r) patient questionnaire for idiopathic scoliosis. *Spine (Phila Pa 1976)* 2013;38:1875-84.

11. Ersberg A, Gerdhem P. Pre- and postoperative quality of life in patients treated for scoliosis. *Acta Orthop* 2013;84:537-43.

12. Cheung KM, Cheng EY, Chan SC, et al. Outcome assessment of bracing in adolescent idiopathic scoliosis by the use of the SRS-22 questionnaire. *Int Orthop* 2007;31:507-11.

13. Ravens-Sieberer U, Wille N, Badia X, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Quality of Life Research* 2010;19:887-97.

14. Perez Sousa MÁ, Olivares Sánchez-Toledo PR, Gusi Fuerte N. Parent-child discrepancy in the assessment of health- related quality of life using the EQ-5D-Y questionnaire. *Arch Argent Pediatr* 2017;115:541-6.

15. Wong CKH, Cheung PWH, Luo N, et al. Responsiveness of EQ-5D Youth version 5-level (EQ-5D-5L-Y) and 3-level (EQ-5D-3L-Y) in Patients With Idiopathic Scoliosis. *Spine (Phila Pa 1976)* 2019;44:1507-14.

16. EQ-5D-3L User Guide [database online]. 2018.

17. Hays RM, Valentine J, Haynes G, et al. The Seattle Pediatric Palliative Care Project: effects on family satisfaction and health-related quality of life. *J Palliat Med* 2006;9:716-28.

18. Hani T. Proxy reporting of quality of life using the EQ-5D. *Medical care* 2002;12.

19. Baker R. Quality of Life Assessment in Clinical Trials: Methods and Practice. *J Rev Soc Med* 1999;92.

20. Fayed N, Avery L, Davis AM, et al. Parent Proxy Discrepancy Groups of Quality of Life in Childhood Epilepsy. *Value in Health* 2019;22:822-8.

21. Wendy J U. Challenges in health state valuation in paediatric economic evaluation: are QALYs contraindicated? *PharmacoEconomics* 2011;8.

22. Shiroiwa T, Fukuda T, Shimozuma K. Psychometric properties of the Japanese version of the EQ-5D-Y by self-report and proxy-report: reliability and construct validity. *Quality of Life Research* 2019.

23. Åström M, Persson C, Lindén-Boström M, et al. Population health status based on the EQ-5D-Y-3L among adolescents in Sweden: Results by sociodemographic factors and self-reported comorbidity. *Quality of Life Research* 2018;27:2859-71.

24. T O. Reported activities of daily living: agreement between elderly subjects with and without dementia and their caregivers. *Age and ageing* 1997;2.

25. Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Quality of Life Research* 2010;19:875-86.
26. Kreimeier S, Greiner W. EQ-5D-Y as a Health-Related Quality of Life Instrument for Children and Adolescents: The Instrument's Characteristics, Development, Current Use, and Challenges of Developing Its Value Set. *Value in Health* 2019;22:31-7.
27. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* 2016;15:155-63.
28. Nahathai W. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC medical research methodology* 2013.
29. Wu XY, Ohinmaa A, Johnson JA, et al. Assessment of children's own health status using visual analogue scale and descriptive system of the EQ-5D-Y: linkage between two systems. *Quality of Life Research* 2014;23:393-402.
30. Wong CKH, Cheung PWH, Luo N, et al. A head-to-head comparison of five-level (EQ-5D-5L-Y) and three-level EQ-5D-Y questionnaires in paediatric patients. *The European Journal of Health Economics* 2019;20:647-56.
31. P J D. Are proxy assessments of health status after stroke with the EuroQol questionnaire feasible, accurate, and unbiased? *Stroke* 1997;10.
32. Alicia D-R. EQ-5D rated by proxy in institutionalized older adults with dementia: psychometric pros and cons. *Geriatrics & gerontology international* 2014;2.
33. A Simon P. Agreement between patient and proxy assessments of health-related quality of life after stroke using the EQ-5D and Health Utilities Index. *Stroke* 2004;2.
34. B J S. Predictors of a child's ability to use a visual analogue scale. *Child: care, health and development* 2003;4.
35. Eiser C, Morse R. Can parents rate their child's health-related quality of life? Results of a systematic review. *Quality of Life Research* 2001;10:347-57.
36. Pablo G-M. Level of agreement between children, parents, and physicians in rating pain intensity in juvenile idiopathic arthritis. *Arthritis and rheumatism* 2006;2.
37. Matsumoto H, G Vitale M, Hyman J, et al. Can parents rate their children's quality of life? Perspectives on pediatric orthopedic outcomes. *Journal of pediatric orthopedics. Part B* 2011;20:184-90.
38. Soleimani F, Vameghi R, Kazemnejad A, et al. Psychometric Properties of the Persian Version of Cerebral Palsy Quality of Life Questionnaire for Children. *Iranian Journal of Child Neurology* 2015;9:76-86.
39. Wandner LD, Scipio CD, Hirsh AT, et al. The perception of pain in others: how gender, race, and age influence pain expectations. *J Pain* 2012;13:220-7.
40. Rafael KM, Hüsseyin S, Rüdiger K. Epidemiology of adolescent idiopathic scoliosis. *Journal of Children's Orthopaedics* 2013;7:3-9.

Figure 1. Distribution (%) of three-level responses to five dimensions (self-report and proxy-report) The majority of patients and proxies reported no problems regardless of any dimensions. There was a tendency that the category of "Some problems" was used more frequently by proxies than patients except for "Doing usual activities" dimension and no patients rated the worst problem level on dimensions (33333).

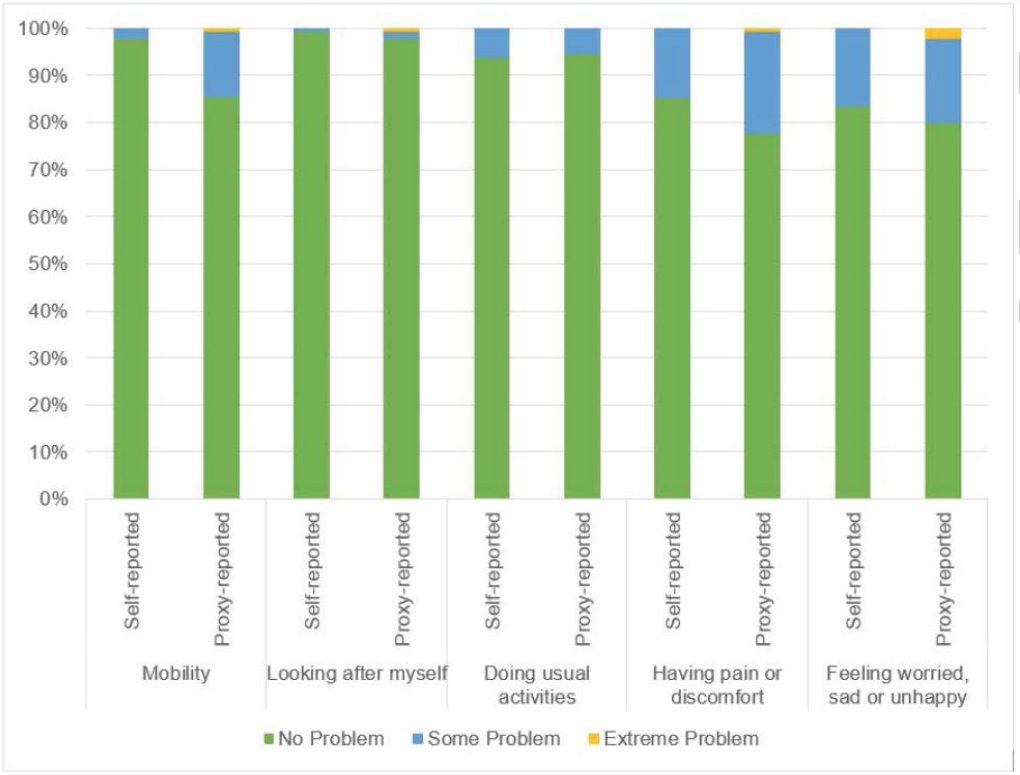

Table 1. Descriptive statistics of EQ-VAS score by characteristics

Characteristics	N (%)	Mean VAS Score (SD)		
Sociodemographic characteristics		*Self-reported*	*Proxy-reported*	*Patient-Proxy Absolute Difference*
Patients				
Gender				
Male	12 (9.5%)	84.9 (11.5)	87.5 (16.1)	10.6 (12.5)
Female	115 (90.6%)	88.8 (9.7)	86.6 (11.7)	8.9 (9.5)
Age (years)				
10-12	21 (16.5%)	90.5 (7.5)	87.7 (8.9)	7.9 (8.7)
13-15	90 (70.9%)	88.1 (10.4)	86.6 (12.6)	9.2 (10.0)
≥16	16 (12.6%)	87.2 (9.3)	85.6 (13.7)	9.4 (10.0)
Body Mass Index (kg/m^2)				
18.5-23.9	50 (40.7%)	86.2 (8.8)	87.7 (12.0)	9.7 (10.0)
<18.5	70 (56.9%)	89.4 (0.5)	85.8 (12.3)	8.7 (9.5)
>23.9	3 (2.4%)	86.7 (5.8)	86.7 (11.6)	6.7 (5.8)
Proxies				
Age (years)				
30-40	33 (28.0%)	——	88.9 (11.4)	8.4 (9.5)
41-50	73 (61.9%)	——	85.2 (12.4)	9.0 (9.9)
51-60	6 (5.1%)	——	90.8 (8.0)	6.2 (4.5)
≥61	6 (5.1%)	——	88.3 (3.7)	13.3 (11.7)
Relationship		——		
Mother	90	——	86.7 (11.8)	9.0 (9.3)

	(71.4%)			
Father	32 (25.4%)	——	86.2 (13.2)	9.5 (11.3)
Grandparent or others	4 (3.2%)	——	93.8 (7.5)	3.8 (4.8)
Education		——		
No education or primary school	9 (7.3%)	——	80.8 (17.0)	12.6 (10.0)
Secondary school	79 (28.5%)	——	87.0 (11.2)	8.3 (8.7)
Tertiary or university	35 (64.2%)	——	87.1 (12.7)	9.0 (11.1)
Living with proxies		——		
No	5 (4.1%)	——	88.0 (12.6)	3.8 (6.1)
Yes	117 (95.9%)	——	86.8 (12.3)	9.3 (9.9)
Patients' clinical characteristics				
Cobb angle, ° (mean ± SD)	25.8 ± 10.0	——	——	——
≤40°, mild or moderate	111 (86.1%)	88.1 (9.9)	86.6 (12.5)	9.1 (9.84)
>40°, severe	9 (7.5%)	92.7 (9.2)	86.1 (9.9)	11.2 (9.67)
Curvature type				
Thoracic curve (type 1/2)	63 (52.9%)	87.4 (10.2)	86.8 (12.1)	9.1 (10.3)
Lumbar curve (type 5)	32 (26.9%)	88.5 (11.0)	86.1 (13.6)	8.6 (9.5)
Thoracic & Lumbar curve (types 3/4/6)	24 (20.2%)	91.1 (7.2)	87.7 (10.5)	9.8 (8.5)
Treatment modality				
Initial presentation by referral	33 (26.8%)	90.3 (10.1)	88.7 (10.7)	7.15 (7.8)
Brace	43 (35.0%)	86.8 (10.9)	88.0 (12.4)	8.74 (10.0)
FU observation	41 (33.3%)	89.4 (8.7)	83.9 (13.6)	11.45 (10.7)

FU observation with bracing before	3 (2.4%)	81.7 (2.9)	83.33 (2.9)		1.67 (2.9)
Before surgery	1 (0.8%)	100	80		20
After surgery return for FU	1 (0.8%)	82	90		8
FU observation + before surgery	1 (0.8%)	100	70		30
Overall score		88.4 (9.9) *	86.7 (12.1) *		9.0 (9.7)

VAS – visual analog scale; SD – standard deviation; FU – follow-up.

*Independent t- test on difference between self-reported and proxy-reported VAS scores, p value was 0.104.

Table 2. Distribution of three-level responses to five dimensions (self-report and proxy-report)

EQ-5D-3L-Y Dimension		Self-reported N	Self-reported %	Proxy-reported N	Proxy-reported %	P-value of Chi-square test
Mobility	No Problems	125	97.7	110	85.3	0.002*
	Some Problems	3	2.3	18	14.0	
	Extreme Problems	0	0.0	1	0.8	
Looking after myself	No Problems	127	99.2	126	97.7	0.513
	Some Problems	1	0.8	2	1.6	
	Extreme Problems	0	0.0	1	0.8	
Doing usual activities	No Problems	120	93.8	122	94.6	0.778
	Some Problems	8	6.3	7	5.4	
	Extreme Problems	0	0.0	0	0.0	
Having pain or discomfort	No Problems	109	85.2	100	77.5	0.211
	Some Problems	19	14.8	28	21.7	
	Extreme Problems	0	0.0	1	0.8	
Feeling worried, sad or unhappy	No Problems	106	83.5	103	79.8	0.210

Some Problems	21	16.5	23	17.8
Extreme Problems	0	0.0	3	2.3

*P value <0.05; **P value<0.001.

Table 3. Proportion of "no problem" responses (ceiling effect) on the self-reported and proxy-reported EQ-5D-3L-Y

Dimension	Self-reported		Proxy-reported		P-value of McNemar test	Reduction in ceiling effects	
	N	%	N	%		Absolute reduction (%)	Relative reduction (%)
Mobility	125	97.7	110	85.3	<0.001**	12.4	12.7
Looking after myself	127	99.2	126	97.7	0.50	1.6	1.6
Doing usual activities	120	93.8	122	94.6	1.00	-0.8	NA
Having pain or discomfort	109	85.2	100	77.5	0.11	7.6	9.0
Feeling worried, sad or unhappy	106	83.5	103	79.8	0.57	3.6	4.3
Full health (11111)	89	69.5	81	62.8	0.16	6.7	9.7

*P value <0.05; **P value<0.001.

Table 4. Patient-proxy agreement using Gwet's AC and percent agreement for EQ-5D-3L-Y dimensions and intra-class correlation coefficients for EQ-VAS score

	Mobility			Looking after myself			Doing usual activities			Having pain or discomfort			Feeling worried, sad or unhappy			VAS Score
Measure	Gwet's AC[a]	PA	P-value[b]	Gwet's AC[a]	PA	P-value[b]	Gwet's AC[a]	PA	P-value[b]	Gwet's AC[a]	PA	P-value[b]	Gwet's AC[a]	PA	P-value[b]	ICC[c]
Overall	0.84	0.85		0.98	0.98		0.94	0.95		0.77	0.81		0.73	0.77		0.29
Subgroups of patient characteristics																
aged 10-12	0.81	0.84	0.86	0.96	0.96	0.22	0.90	0.91	0.41	0.70	0.79	0.11	0.67	0.73	0.49	0.03
aged 13-16	0.85	0.86		1.00	1.00		0.97	0.97		0.78	0.82		0.78	0.81		0.31
Cobb angle ≤40°	0.82	0.84	0.26	0.98	0.98	0.58	0.93	0.94	0.29	0.75	0.79	0.39	0.76	0.79	0.14	0.31
Cobb angle >40°	0.93	0.94		1.00	1.00		1.00	1.00		0.84	0.88		0.54	0.63		NA

treatment modality - observation	0.83	0.85	0.82	1.00	1.00	0.06	0.98	0.98	0.04*	0.83	0.87	0.02*	0.76	0.80	0.29					0.29
treatment modality - bracing	0.82	0.86		0.96	0.96		0.87	0.90		0.64	0.71		0.65	0.71						0.35
Curvature type - Thoracic	0.84	0.86		0.97	0.97		0.97	0.97		0.67	0.78		0.74	0.77						0.24
Curvature type - Lumbar curve	0.85	0.88	0.67	1.00	1.00	0.41	0.98	0.98	0.04*	0.87	0.88	0.46	0.77	0.80	0.27					0.48
Curvature type - Thoracic & Lumbar	0.80	0.82		1.00	1.00		0.85	0.88		0.81	0.88		0.57	0.66						0.01
Subgroups of proxy characteristics																				
mother	0.83	0.85	0.83	0.99	0.99	0.49	0.92	0.93	0.40	0.73	0.81	0.92	0.70	0.75	0.33					0.29
father, grandparent or others	0.83	0.87		0.97	0.97		0.97	0.97		0.74	0.78		0.80	0.83						0.31
aged <40	0.77	0.81	0.48	1.00	1.00	0.47	0.92	0.92	0.58	0.77	0.81	0.97	0.67	0.73	0.58					0.18

aged ≥40	0.84	0.85		0.98	0.98		0.94	0.95		0.77	0.81		0.74	0.77		0.32
education - secondary or below	0.78	0.80	0.02*	0.99	0.99	0.49	0.92	0.93	0.40	0.77	0.84	0.20	0.71	0.75	0.63	0.36
education - tertiary or university	0.97	0.97		0.97	0.97		0.97	0.97		0.70	0.74		0.76	0.79		0.22

Gwet's AC - Gwet's agreement coefficient; PA – perfect agreement; ICC - intra-class correlation coefficients.

Notes:

[a]The Gwet's AC-statistic is a measure of the reliability of agreement. The higher the Gwet's AC, the higher the amount of agreement reliability. Gwet's AC of <0.2 was interpreted as poor agreement of individual domain responses between self-reported and proxy-reported assessments; 0.21-0.4 as fair; 0.41-0.6 as moderate; 0.61-0.8 as good and ≥0.8 as very good.

[b]P-value of Chi-squared test on significant difference between subgroups and agreement of responses of each dimension.

[c]The intra-class correlation coefficients (ICC) is a measure of the reliability of agreement. The higher the ICC, the higher the amount of agreement reliability. ICC of <0.4 was interpreted as poor agreement of EQ-VAS score between two assessments; 0.4-0.74 as moderate; >=0.75 as good.

*P value <0.05.

Table 5. Regression analysis on agreement of EQ-VAS score and dimension responses between self-reported and proxy-reported EQ-5D-3L-Y

Regression model	Linear regression			Logistic regression			
Independent Variables	Coefficient (95% CI)			Odds Ratio (95% CI)			
	Self-reported VAS Score	Proxy-reported VAS Score	Patient-Proxy VAS Score Absolute Difference	Mobility	Doing usual activities	Having pain or discomfort	Feeling worried, sad or unhappy
Constant	102.0 (80.1,124.0)	89.7 (58.2,121.3)	2.0 (-23.2,27.2)	11.8 (0.4,396.5)	11.3 (0.9,145.4)	7.0 (0.2,206.6)	2.8 (0.1,81.4)
Gender (patient)							
Male	reference	reference	reference	reference	reference	reference	reference
Female	4.1 (-2.4,10.6)	-7.0 (-16.3,2.2)	3.3 (-4.1,10.7)	0.4 (0.0,4.7)	1.0[a]	0.6 (0.1,5.8)	0.7 (0.1,7.0)
Age (years, patient)	-0.8 (-2.2,0.5)	-0.6 (-2.4,1.3)	0.1 (-1.3,1.6)				
10-12				reference	reference	reference	reference
≥13				1.8 (0.6,5.7)	3.0 (0.4,22.1)	1.1 (0.4,3.1)	1.5 (0.5,4.0)
Body Mass Index (patient)	-0.3 (-1.1,0.4)	0.7 (-0.3,1.7)	-0.2 (-1.0,0.6)				
18.5-23.9				reference	reference	reference	reference
<18.5 or >23.9				0.8 (0.2,2.7)	1.3 (0.2,8.2)	2.0 (0.7,5.8)	0.5 (0.2,1.4)
Curvature type (patient)							
Thoracic curve (type 1/2)	reference	reference	reference	reference	reference	reference	reference
Lumbar curve (type 5)	0.7 (-3.8,5.2)	-0.4 (-6.4,5.5)	0.0 (-4.8,4.7)	1.3 (0.3,5.5)	0.7 (0.1,9.5)	1.2 (0.3,4.3)	1.6 (0.4,6.2)
Thoracic & Lumbar curve (types 3/4/6)	3.5 (-1.2,8.3)	2.2 (-4.4,8.7)	1.0 (-4.2,6.2)	0.7 (0.2,3.0)	0.1* (0.0,0.7)	2.4 (0.5,10.8)	0.4 (0.1,1.4)

Cobb angle (°, patient)							
≤40°, mild or moderate	reference	reference	reference	reference	reference	reference	reference
>40°, severe	2.3 (-5.0,9.7)	2.1 (-7.7,11.9)	-1.9 (-9.7,5.9)	1.0[a]	1.0[a]	0.6 (0.1,3.9)	0.6 (0.1,3.7)
Treatment modality (patient)							
observation	reference	reference	reference	reference	reference	reference	reference
bracing	-3.3 (-7.3,0.6)	2.2 (-3.2,7.6)	-1.7 (-6.0,2.6)	2.3 (0.6,8.6)	0.6 (0.1,5.2)	0.3* (0.1,0.9)	0.8 (0.3,2.3)
Age (years, proxies)		0.0 (-0.4,0.3)	0.0 (-0.3,0.3)				
30-40				reference	reference	reference	reference
41-50				1.2 (0.3,4.3)	1.6 (0.2,11.4)	0.9 (0.3,3.4)	1.2 (0.4,3.8)
≥51				3.3 (0.2,43.7)	1.0[a]	0.9 (0.1,7.9)	0.3 (0.1,1.9)
relationship with patients (proxies)							
Mother		reference	reference	reference	reference	reference	reference
Father/Grandparent or others		-1.8 (-8.2,4.6)	2.4 (-2.7,7.5)	0.4 (0.1,1.8)	1.7 (0.1,26.7)	0.6 (0.2,2.2)	3.2 (0.78,12.9)
Education (proxies)							
No education/Primary or secondary school		reference	reference	reference	reference	reference	reference
Tertiary/university		-0.3 (-5.9,5.3)	1.3 (-3.1,5.8)	9.4* (1.0,87.2)	1.2 (0.1,15.5)	0.4 (0.1,1.3)	0.9 (0.3,3.1)
Living with patients							

(proxies)							
No		reference	reference	reference	reference	reference	reference
Yes		-1.5 (-13.8,10.8)	6.1 (-3.7,16.0)	0.4 (0.0,5.9)	1.0[a]	1.5 (0.1,18.4)	2.3 (0.2,29.2)
R-squared/Log likelihood	0.09	0.06	0.05	-40.14	-18.51	-49.05	-52.18

CI – confidence intervals.

[a]Odds ratio of 1 represents that the odds of the responses of the dimension given the presence of particular subgroup is equal to the odds of the responses of that dimension occurring in the absence of the reference subgroup, indicating that no association was observed between the subgroup and the dimension problem levels.

*P value <0.05; **P value<0.001.