Repeatability, Reliability, and Concurrent Validity of the Scoliosis Research Society-22 Questionnaire and EuroQol in Patients With Adolescent Idiopathic Scoliosis

Raphael D. Adobor, MD,* Silje Rimeslåtten, RN,* Anne Keller, PhD, MD,† and Jens Ivar Brox, PhD, MD*

Most idiopathic scoliosis cases are of the late-onset type, classified as adolescent idiopathic scoliosis (AIS). Bracing and surgery are aimed at preventing curve progression.

Study Design. Cross-sectional study in patients with adolescent idiopathic scoliosis (AIS).

Objectives. To evaluate the repeatability, reliability, internal consistency, and concurrent validity (CD) of an adapted Norwegian version of the Scoliosis Research Society 22 questionnaire (SRS-22) and the generic health-related quality of life instrument EuroQol (EQ-5D and EQ-VAS).

Summary of Background Data. SRS-22 is widely used for evaluation of health-related quality of life in AIS. Its repeatability, which is essential for use in follow-up studies, and CD with EuroQol which can be used for cost-utility analysis, has not yet been assessed.

Methods. The forward-backward translation of the English version of the SRS-22 was performed according to guidelines for cross-cultural adaptation of outcome questionnaires. Fifty-seven patients of various ages with AIS and deformity severity filled out standardized questionnaires: SRS-22, EQ-5D, and EQ-VAS, each twice in a 2-week interval. The study was approved by the Regional Ethics Committee for Medical Research in Norway.

Results. There were no floor or ceiling effects on the score distributions. The study demonstrated moderate internal consistency and high reliability of SRS-22 questionnaire with Chronbach α and intraclass correlation coefficient ranging from 0.76 to 0.93 for the 5 domains. Repeatability was excellent for all SRS-22 domains with repeatability coefficients <1. CD with EQ-5D was poor to moderate with Pearson’s r ranging from 0.14 to 0.58. However, total scores of the 2 instruments showed satisfactory agreement.

Conclusion. The SRS-22 outcome instrument has satisfactory repeatability, but CD with EQ-5D suggests that the disease-specific and the generic questionnaire measure different constructs.

Key words: adolescent idiopathic scoliosis, SRS-22, EuroQol, repeatability. Spine 2010;35:206–209

Previously, patients with AIS were commonly monitored by clinical evaluation and objective radiologic measures. Today, clinical outcome research focuses on outcome from the patient’s perspective in addition to evaluations of care providers. The Scoliosis Research Society (SRS) has developed a simple, practical, disease-specific, patient-based assessment for AIS. The SRS-22 questionnaire is currently accepted internationally for assessment of health-related quality of life for AIS. The SRS-22 has recently been translated and validated in Spanish, Turkish, Japanese, and Chinese versions.

Developers of the original SRS-22 questionnaire and subsequent validators have limited the assessment of reproducibility to calculation of intraclass correlation coefficient (ICC). This statistical measure of reliability describes the ability to discriminate between individuals, but agreement parameters are required for evaluation of measurement error in follow-ups.

The statistical measure proposed by Altman and Bland estimates repeatability, which is a measure based on the variation within individuals. The SRS-22 has been concurrently validated against the Medical Outcome Study Short Form-36, which is a generic questionnaire measuring health-related quality of life. The EuroQol (EQ-5D and EQ-VAS) is a short form generic health-related quality of life questionnaire that has been used for patients with back pain. The SRS-22 has not yet been concurrently validated against the EuroQol.

Our main objective was to evaluate the agreement and concurrent validity (CD) of the SRS-22 and EuroQol questionnaires in patients with AIS.

Materials and Methods

The Translation Process

The English version of the SRS-22 was translated into Norwegian according to routines for cross-cultural adaptation of outcome questionnaires. First, the English version of the SRS-22 was translated into Norwegian and retranslated back to English by 2 independent bilingual translators—one whose mother tongue is English and the other whose mother tongue is Norwegian. A review committee composed of 1 specialist in physical medicine and rehabilitation, 1 public health nurse, 2 spine surgeons, and the 2 translators further assessed the forward and backward translations, and a consensus was achieved on the final translation.

Patients

The final version of the SRS-22, and the EQ-5D, and EQ-VAS questionnaires were mailed to AIS patients of various ages, and...
Repeatability, Reliability, and Concurrent Validity • Adobor et al 207

severity of curves treated at Rikshospitalet University Hospital, Oslo, Norway. Seventy-six patients answered the questionnaire the first time and 57 patients (75%) answered the second time. Results from these 57 were included in the study. There were 9 (16%) men and 48 (84%) women. Median age at the time of investigation was 21 years (range: 12–45 years). Median Cobb angle at the time of investigation was 37° (range: 12°–71°). Twelve (21%) had previously been treated with a brace, 6 (11%) were currently braced, 22 (39%) had surgery, and 17 (30%) were scheduled for surgery.

**Questionnaires**

The SRS-22 covers 5 domains—4 with 5 questions each: function/activity, pain, self-perceived image, and mental health; and one with 2 questions: satisfaction with treatment. Each item has 5 verbal response alternatives ranging from 1 (worst) to 5 (best). Consequently, each domain has a total sum score ranging from 5 to 25, except for satisfaction, which ranges from 2 to 10. The sum of the first 4 domains gives a maximum subtotal of 100, and with addition of the satisfaction domain a maximum total sum of 110. Results are usually expressed as the mean (total sum of the domain divided by the number of items answered) for each domain.

The EQ-5D questionnaire evaluates 5 dimensions: mobility, self-care, activities of daily life, pain and anxiety/depression. Each dimension is described by 3 possible levels of problem (none, moderate, and severe). Scores are transformed using utility weights from the general population to produce a single index, ranging from –0.59 for the worst possible health state to 1.00 for the best possible health state. EQ-VAS is a vertical visual analogue scale presented as a thermometer on which the patients rate their overall current health from 0 (worst imaginable health) to 100 (best imaginable health).

A test-retest design was used. Patients answered the questionnaire a second time after 2 weeks, and it was assumed that their condition was unchanged. All patients received oral and written information about the project and gave their informed consent. The study was approved by the Regional Ethics Committee for Medical Research in Norway.

**Statistical Analysis**

Means, standard deviations, and frequencies were calculated for each domain. Two measures of reliability were estimated: internal consistency (Chronbach α) and Intraclass correlation coefficient (ICC; 2,1). The repeatability coefficient was estimated as suggested by Bland and Altman.12,13 The measurement of repeatability is based on the standard error of measurement (SEM) which is calculated by extracting the square root of the mean square root within subject variance term in the one-way analysis of variance table. The coefficient of repeatability (CR) is then calculated from the formula: CR = SEM × 1, 96√2. The difference between 2 measurements for the same subject is suspected to be less than the repeatability coefficient for 95% of pairs of observations. This measure defines the smallest detectable change between 2 measurements on the same individual and has been designated the minimal detectable change.2 As recommended by Bland and Altman, plots of the difference between test and retest against the mean of the sum scores were constructed for detecting any evidence of increasing variability with higher mean scores (heteroscedasticity). Two SD of the difference were subtracted or added to the mean difference to create “limits of agreement” which were drawn as lines in the plots. CD comparing SRS-22 with EQ-5D questionnaire was estimated using Pearson’s correlation coefficient (r) in numerical scales. The Statistical Package for Social Science (SPSS), version 14.0 (SPSS Inc., Chicago, IL), was used to analyze the data except for Altman plots which were constructed using MedCalc 9.

**Results**

The distribution of scores for the 5 SRS-22 domains is shown in Table 1. The mean total score for all patients was 3.9 (SD: 0.6). There were low levels of floor and ceiling effects. Self-image and satisfaction with treatment domains recorded a floor score. The maximum average domain score of 5 was reached for pain, mental health, and satisfaction with treatment domains for less than 10% of the patients. Internal consistency reliability (Cronbach α) was higher than 0.85 for all domains (Table 2). ICC was higher than 0.80 for all domains except for the function/activity domain (0.76) (Table 2). The CR for each SRS-22 domain was equal to or less than 1 point and acceptable for the EQ-5 index and EQ-VAS (Table 2). The CD for SRS domains with EuroQol is shown in Table 3. Figure 1 shows plots of the difference between the answers from the first and second time for SRS-22 and EQ-5D and EQ-VAS.

**Discussion**

The present study demonstrates satisfactory score distribution, internal consistency, reliability, and repeatability.

**Table 1. Description Statistics on Individual Domain Scores**

| Domain               | Mean (SD) | Floor Score | Ceiling Score | Floor Score 1 (%) | Ceiling Score 5 (%) |
|----------------------|-----------|-------------|---------------|-------------------|---------------------|
| SRS-22 (0–5)         |           |             |               |                   |                     |
| Pain                 | 4.0 (0.79)| 2.0         | 5.0           | 0                 | 10.5                |
| Self-image           | 3.5 (0.85)| 1.0         | 4.8           | 1.8               | 0                   |
| Function/activity    | 4.1 (0.57)| 2.4         | 4.6           | 0                 | 0                   |
| Mental health        | 3.9 (0.68)| 2.2         | 5.0           | 0                 | 10.0                |
| Satisfaction with treatment | 4.0 (0.99) | 1.0         | 5.0           | 4.2               | 7.3                 |
| EuroQol              |           |             |               |                   |                     |
| EQ-SD (−0.49–1.00)   | 0.83 (0.12)| −0.77       | 1.0           | 1.8               | 35                  |
| EQ-VAS (0–100)       | 83 (15.5) | 20          | 100           | 1.8               | 4.7                 |

**Table 2. Internal Consistency (Cronbach α), Intraclass Coefficient (ICC), and Coefficient of Repeatability (CR) (N = 57)**

| Domain               | Cronbach α (95% CI) | ICC 2,1 (95% CI) | Repeatability Coefficient |
|----------------------|----------------------|-----------------|---------------------------|
| SRS-22 (1–5)         |                      |                 |                           |
| Pain                 | 0.93 (0.88–0.96)     | 0.87 (0.79–0.92)| 0.78                      |
| Self-image/ appearance | 0.93 (0.88–0.96)     | 0.87 (0.78–0.92)| 0.98                      |
| Function/activity    | 0.87 (0.77–0.92)     | 0.76 (0.63–0.85)| 0.74                      |
| Mental health        | 0.89 (0.82–0.94)     | 0.80 (0.69–0.88)| 0.82                      |
| Satisfaction with treatment | 0.90 (0.77–0.96) | 0.82 (0.63–0.92) | 1.08                      |
| EuroQol              |                      |                 |                           |
| EQ-SD (−0.59–1.00)   | 0.89 (0.81–0.94)     | 0.80 (0.68–0.88)| 0.25                      |
| EQ-VAS (0–100)       | 0.95 (0.92–0.97)     | 0.91 (0.85–0.95)| 13.05                     |

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
Table 3. Concurrent Validity of the SRS-22 Domains With Relevant EQ-5D Profiles as Determined by Pearson's Correlation Coefficient $r$ (n = 57)

| SRS-22 Domain       | Euro Qol       | $r$    | CI       | $P$  |
|---------------------|----------------|--------|----------|------|
| Pain                | Pain           | -0.49  | -0.67 to -0.26 | <0.00|
|                     | EQ-5D          | 0.59   | 0.38–0.74 | <0.00|
|                     | EQ-VAS         | 0.40   | 0.15–0.60 | <0.00|
| Self-image          | Anxiety/depression | -0.42 | 0.17–0.61 | <0.00|
|                     | EQ-5D          | 0.62   | 0.42–0.76 | <0.00|
|                     | EQ-VAS         | 0.44   | 0.12–0.63 | <0.00|
| Function/activity    | Mobility       | -0.17  | -0.42–0.21 | 0.21 |
|                     | Self-care      | 0.43   | 0.18–0.62 | <0.00|
|                     | Usual activities | -0.14 | 0.13–0.38 | 0.31 |
|                     | EQ-5D          | 0.36   | 0.10–0.57 | <0.00|
|                     | EQ-VAS         | 0.52   | 0.30–0.69 | <0.00|
| Mental health       | EQ-5D          | 0.57   | 0.35–0.72 | <0.00|
|                     | EQ-VAS         | 0.35   | 0.10–0.56 | <0.00|
| Subtotal mean score | EQ-5D          | 0.67   | 0.49–0.79 | <0.00|
| Subtotal mean score | EQ-VAS         | 0.57   | 0.36–0.73 | <0.00|

We found no reports of repeatability measurements of the SRS 22 in the literature. The SEM quantifies the precision of individual scores on a test, and is therefore advocated in conjunction with ICC. The coefficients of repeatability for the 5 domains of pain, self-image, function/activity, mental health, and satisfaction with treatment, varied from 0.74 to 1.00. The practical interpretation is that the lowest detectable average change for each domain is 1 step on the 5 level verbal scales with response categories given labels from 1 to 5. A smaller change between 2 subsequent measurements is indistinguishable from the measurement error, and the given limit represents the minimum detectable difference. The variation between 2 measurements in the same individual should be taken into consideration when assessing for all domains of SRS-22 and for EQ-5D and EQ-VAS. While reliability parameters are recommended for instruments that are used for discriminative purposes, agreement parameters are required for use in follow-up studies. The ICC mainly refers to the variability between subjects in the population sample, and the trial-to-trial noise within subjects usually has minor influence on the calculated values. The ICCs in the present study are in agreement with previous studies and support the conclusion that SRS-22 is a useful instrument for discrimination between patients.

![Figure 1. Altman plots showing intraindividual differences between tests plotted against the means for SRS-domains and sub-total scores, EQ-5D, and EQ-VAS. Means and 95% limits of agreement are given.](image-url)
follow-up results after treatment and in the planning of prospective studies. Repeated measurements may reduce measurement error and increase the validity of observations. The estimates of both agreement and reliability parameters reported in this study supplement current knowledge of clinimetric properties of questionnaires for use in patients with AIS. The study was undertaken in a clinical setting, including an adequately sized population of patients and presuming that the condition was unchanged between questionnaire administrations.

Internal consistency for the function domain was slightly lower than reported in the original version and subsequent Spanish and Turkish transcultural adaptations in a younger age group.1,5,7 The inconsistencies have previously been traced to question 15 (Are you and/or your family experiencing financial difficulties because of your back?) and question 18 (Does your back condition limit your going out with friends/family?). These questions may reflect social aspects (economy and participation) which may differ from function in terms of the ability to perform activities of daily living. There is also a perception that question 15 might not be applicable for countries with a public health care system as in Norway. However, the similar mean overall Cronbach α compared with the original American version reflects similarity of the 2 cultures. In the current study, Cronbach α values for questions 15 and 18 of the function domain were 0.73 and 0.77, which compare well with the original and the translated forms of the SRS-22 questionnaire in similar patient groups. Because of this, we suggest retaining questions 15 and 18. The SRS-22 has recently been modified and refined, resulting in a higher Cronbach α value for the function domain.3

We found poor correlations between SRS-22 and EQ-5D for the main dimensions of pain, mobility, function, and mental health. The self-image and satisfaction to treatment domains are not comparable between the 2 instruments. The overall scores of the 2 instruments showed moderate correlations. Employing the repeatability measure as proposed by Altman and Bland showed acceptable agreement (Figure 1). The low correlation coefficient observed in the present study may reflect poor intrinsic correlation between the 2 instruments rather than a validity problem of the translated questionnaire. Another possible reason for the low correlation coefficient is that the EuroQol has been validated for use in adult populations with back pain, but not in the younger population with spine deformity as in our study population. The advantage of the EuroQol is that it can be used to compare with other diseases and to achieve a utility index in cost effectiveness evaluation.

For further use, EuroQol’s responsiveness as compared with SRS-22 warrants further assessment. In summary, we found satisfactory reproducibility of the SRS-22 for use in follow-up studies, but poor CD compared with the generic EuroQol instrument.

**Key Points**

- The Norwegian version of the SRS-22 and EuroQol were evaluated in 57 patients with AIS.
- Repeatability, internal consistency, and reliability were acceptable.
- The poor validity compared with EuroQol support the use of a specific questionnaire for assessment of AIS.

**References**

1. Alanay A, Çil A, Berk H, et al. Reliability and validity of adapted Turkish Version of Scoliosis Research Society-22 (SRS-22) questionnaire. Spine 2005;30:2464–8.
2. Altman DG. Practical Statistics for Medical Research. London, United Kingdom: Chapman and Hall; 1991.
3. Asher M, Min LS, Burton D, et al. The reliability and concurrent validity of the scoliosis research society-22 patient questionnaire for idiopathic scoliosis. Spine 2003;28:63–9.
4. Asher MA, Lai SM, Glattes RC, et al. Refinement of the SRS-22 health-related quality of life questionnaire function domain. Spine 2006;31:593–7.
5. Bago J, Climent JM, Ey A, et al. The Spanish version of the SRS-22 patient questionnaire for idiopathic scoliosis: transcultural adaptation and reliability analysis. Spine 2004;29:1676–80.
6. Cheung KM, Senkoylu A, Alanay A, et al. Reliability and concurrent validity of the adapted Chinese version of Scoliosis Research Society-22 (SRS-22) questionnaire. Spine 2007;32:1141–5.
7. Climent JM, Bago J, Ey A, et al. Validity of the Spanish version of the Scoliosis Research Society-22 (SRS-22) patient questionnaire. Spine 2005;30:703–9.
8. Hashimoto H. Validation of a Japanese version of the Scoliosis Research Society-22 patient questionnaire among idiopathic scoliosis patients in Japan. Spine 2007;32:E141–6.
9. de Vet HC, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033–9.
10. Solberg TK, Olsen JA, Ingebrigtsen T, et al. Health-related quality of life assessment by the EuroQol-5D can provide cost-utility data in the field of low-back surgery. Eur Spine J 2005;14:1000–7.
11. Guillen LM, Cross-cultural adaptation and validation of health status measures. Scand J Rheumatol 1995;24:61–3.
12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
13. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60.