North American Spine Society Journal (NASSJ) 19 (2024) 100545

Contents lists available at ScienceDirect

North American Spine Society Journal (NASSJ)

journal homepage: www.elsevier.com/locate/xnsj

Clinical Studies

Psychometric analysis and the implications for the use of the scoliosis research society questionnaire (SRS-22r English) for individuals with adolescent idiopathic scoliosis

Donna J. Oeffinger, PhD[a,*], Henry Iwinski, MD[a], Vishwas Talwalkar, MD[a], David M. Dueber, PhD[b]

[a] Shriners Children's Lexington, Lexington, KY, 110 Conn Terrace, Lexington, KY 40508, United States
[b] The Herb Innovation Center, University of Toledo, 3100 Gillham Hall, Toledo, OH 43606, United States

ARTICLE INFO

Keywords:
SRS-22r (Scoliosis Research Society Questionnaire-22r)
AIS (Adolescent Idiopathic Scoliosis)
Psychometric analysis
Item response theory (IRT)
Structural equation modeling (ESEM)
Patient reported outcomes (PROs)

ABSTRACT

Background: Despite widespread usage of the SRS-22r questionnaire (Scoliosis Research Society Questionnaire-22r), the English version has only sparingly been subjected to analysis using modern psychometric techniques for patients with adolescent idiopathic scoliosis (AIS). The study purpose was to improve interpretation and clinical utility of the SRS-22r for adolescents with AIS by generating additional robust evidence, using modern statistical techniques. Questions about (1) Structure and (2) Item and Scale Functioning are addressed and interpreted for clinicians and researchers.

Methods: This retrospective case review analyzed SRS-22r data collected from 1823 patients (mean age 14.9±2.2years) with a primary diagnosis of AIS who clinically completed an SRS-22r questionnaire.

Individual SRS-22r questions and domain scores were retrieved through data queries. Patient information collected through chart review included diagnosis, age at assessment, sex, race and radiographic parameters. From 6044 SRS-22r assessments, 1 assessment per patient was randomly selected. Exploratory structural equation modeling (ESEM) and item response theory (IRT) techniques were used for data modeling, item calibration, and reliability assessment.

Results: ESEM demonstrated acceptable fit to the data: χ^2 (130)=343.73, p<.001; RMSEA=0.035; CFI=0.98; TLI=0.96; SRMR=0.02. Several items failed to adequately load onto their assigned factor. Item fit was adequate for all items except SRSq10 (Self-Image), SRSq16 (Mental Health), and SRSq20 (Mental Health). IRT models found item discriminations are within normal levels for items in psychological measures, except items SRSq1 (pain), SRSq2 (pain), and SRSq16 (mental health). Estimated reliability of the Function domain (ρ=0.69) was low, however, Pain, Self-Image and Mental Health domains exhibited high (ρ>0.80) reliability.

Conclusions: Modern psychometric assessment of the SRS-22r, in adolescent patients with AIS, are presented and interpreted to assist clinicians and researchers in understanding its strengths and limitations. Overall, the SRS-22r demonstrated good psychometric properties in all domains except function. Cautious interpretation of the total score is suggested, as it does not reflect a single HRQoL construct.

Introduction

Adolescent idiopathic scoliosis (AIS), is among the most common spinal deformities, occurring in as many as 4% of children between the ages of 10–18years [1]. Reflecting the larger national trend, patient-centered patient-reported outcome questionnaires (PROs) have been adopted in clinical management of patients with AIS [2,3]. Used clinically, PROs provide information about a patient's health related quality of life (HRQoL), facilitate patient-provider communication, and provide objective documentation of changes and treatment outcomes. PROs used in research assess treatment effectiveness and allow for comparisons across studies and institutions [4].

Condition-specific PROs can be more sensitive to change and better reflect the patient's symptoms and functioning [5]. For AIS, the Scoliosis Research Society Questionnaire (SRS-24), developed by Haher et al., [6,7] as a disease-specific HRQoL measure for those undergoing poste-

FDA device/drug status: Not applicable.
Author disclosures: ***DJO:*** Nothing to disclose. ***HI:*** Nothing to disclose. ***VT:*** Nothing to disclose. ***DMD:*** Nothing to disclose.

* Corresponding author. Shriners Children's Lexington, 110 Conn Terrace, Lexington, KY 40508, USA.
E-mail address: doeffinger@shrinenet.org (D.J. Oeffinger).

https://doi.org/10.1016/j.xnsj.2024.100545
Received 10 May 2024; Received in revised form 26 July 2024; Accepted 27 July 2024
Available online 31 July 2024

rior spinal fusion was validated in adults and included mental health questions designed for adults from SF-36. This version, studied extensively, underwent several modifications (SRS-22r, SRS-23, SRS-24, SRS-30) [6,8–10]. Currently, the Scoliosis Research Society (SRS) states that "all practitioners should utilize the SRS-22r, and its various translations, as it is the most recent and most validated version of the SRS Patient Based Outcomes Questionnaire" [11]. The SRS-22r assesses 5 domains: pain, self-image/appearance, function/activity, mental health, and satisfaction with treatment.

The SRS-22r is one of the most widely used and accepted outcome measures assessing HRQoL in patients with scoliosis and has been translated and independently validated in 11 other languages [4,10,12–28]. Traditional psychometric properties of validity, reliability, and responsiveness have been widely studied. Reliability of sub-domain scores has been found to be adequate [29,30], although some studies in other languages found questionable reliability for the function domain [14,16,18,19,21]. Convergent validity in AIS samples, in the form of correlations with similar domains from other HRQoL instruments, has been repeatedly tested; the function domain correlates modestly with function domains of other instruments [13,14,26,28–30], the pain domain correlates excellently with pain domains of other instruments [13,14,26,28–30], and the mental health domain correlates modestly [13,28,30] to excellently [14,26,29,30] with mental health domains of other instruments. No corresponding domains for the self-image and satisfaction domains were present in the other HRQoL instruments employed in these studies. Discriminant validity in AIS samples, in the form of expectedly low correlations with unrelated variables (e.g., between the mental health domain and maximum Cobb angle), has been consistently found [13,15,26,28–32].

Despite the widespread usage and SRS recommendations to include SRS-22r data in all scoliosis research, the English version has only sparingly been subjected to analysis using modern psychometric techniques such as Rasch analysis, item response theory, and factor analysis for adolescent patients [31,32]. These techniques are necessary to assess assumptions tacitly made about items and their functioning when computing and interpreting domain scores for clinical and research purposes. A recent study by Alamrani et al., using qualitative techniques, questions the appropriateness of some individual SRS-22r items when used in the pediatric AIS population [6,33].

The purpose of this study was to improve interpretation and clinical utility of the SRS-22r for adolescents with AIS by generating additional robust evidence, using modern statistical techniques, to assist clinicians in understanding to what degree they can confidently interpret SRS-22r total and individual domain subscale scores in this population. Specifically, the study question that addressed Structure was: Does the hypothesized model with 5 distinct yet related domains, fit the data collected from the SRS-22r? To address Item and Scale Functioning the questions asked were: Are individual SRS-22r items effective for measuring the domains to which they belong? And Are SRS-22r domain scores reliable? The answers to these questions are explored and interpreted for the clinician and researcher. Since the project's primary focus is to report the results of the stringent psychometric analyses and to translate the implications of these findings for clinicians and researchers when using the SRS-22r, only results of the psychometric analyses supporting the clinical and research interpretations are reported. Refer to the Measurement Appendix A for additional details of the statistical analyses and research questions related to the data structure.

Material and methods

Data from individuals with a primary diagnosis of AIS, who clinically completed at least 1 SRS-22r assessment between March 2015 and September 2021 were analyzed. After IRB approval, individual SRS-22r question (item) and overall domain scores, and assessment dates were retrieved from outcome databases and electronic medical record (EMR) queries. When available, PROMIS Pain Intensity (PROMIS Numeric Rating Scalev1.0–Pediatric Pain Intensity1a) scores were collected.

The study population generated from a consecutive case review included all eligible male and female patients and all ethnic backgrounds. Data from those with secondary diagnoses including, but not limited to: cerebral palsy, spina bifida, neuromuscular scoliosis, juvenile idiopathic scoliosis, or other medical diagnoses that could influence quality of life and/or SRS-22r answers were excluded. Patient information collected through EMR review included diagnosis, age at assessment, sex, race and radiographic parameters. From 6,044 assessments, 1 assessment per patient was randomly selected, resulting in 1,823 assessments included in the analyses. Additionally, 206 patients completed the PROMIS pain intensity questionnaire on the same date.

Data analysis plan

Exploratory structural equation modeling (ESEM) and item response theory (IRT) techniques were used for data modeling, item calibration, and reliability assessment. Invariance of item functioning across sex was assessed using ordinal logistic regression and IRT techniques were employed to assess the psychometric properties. ESEM subsumes and expands on the traditional exploratory factor analysis framework for investigating the internal structure of multidimensional data; models fit with ESEM can be interpreted in a confirmatory sense, enable inspection of correlated uniqueness, can be used for further analysis of the latent variables created, and generally enjoy nearly all the features and advantages of structural equation modelling [34]. Likewise, IRT offers a number of advantages over traditional approaches to item analysis and the assessment of measurement precision, including more detailed and robust information about each item's performance, an assessment of measurement precision that is specific to an individual's score rather a single estimate of reliability such as Cronbach's α, and a robust framework for evaluating the extent to which a measure's scores are interpretable in a common way across groups [35,36].

Preliminary data analysis consisted of creating correlation matrices for SRS-22r items and subdomains. Various item and subdomain properties were calculated including means, standard deviations, and proportion of sample at the floor or ceiling of possible values. Internal consistency of the subdomains was estimated using Cronbach's alpha.

The proposed structure of the SRS-22r, consisting of 5 related but distinct subdomains, was tested using ESEM. Fitting of the ESEM was performed in Mplus 8.7 [37] using maximum likelihood estimation with robust standard errors (MLR) and target rotation. ESEM differs from traditional confirmatory factor analysis (CFA) techniques in that crossloadings are allowed, subject to identification constraints [38].

Exact model fit was determined using a standard chi-square goodness of fit test. In the event that exact fit was not found, approximate fit was evaluated based on common fit index criteria: root mean square error of approximation (RMSEA<0.05) [39]; comparative fit index (CFI>0.95) and Tucker Lewis Index (TLI>0.9)5 [40]; and standardized root mean square residual (SRMR<0.05) [41]. In order for model fit to be deemed acceptable, all residual correlations should be less than 0.100 [42]. Items are considered to appropriately belong to their subdomain if they load adequately (standardized factor loading >0.40) on the appropriate factor and do not have a strong (>0.30) crossloading on another factor [43]. Because crossloadings can create bias in the estimates of correlation between observed subdomain scores, [34] inter-factor correlations from the ESEM were descriptively compared to observed correlations between subdomains. Noting that correlations computed within an ESEM are disattentuated, these correlations are expected to be somewhat higher than correlations between observed scores.

Item and scale functioning were assessed using IRT techniques. For each SRS domain, a separate unidimensional graded response IRT model was fit. Item fit was checked using S-χ^2 [44] with a conservative probability threshold of 0.01; empirical item fit plots were examined for items flagged for poor fit. The assumption of local independence was

Table 1
Correlations for SRS-22r subdomains and descriptive data.

	SRS-22r Function	SRS-22r Pain	SRS-22r Self-Image	SRS-22r Mental Health	SRS-22r Satisfaction
SRS-22r Function	(0.69)				
SRS-22r Pain	0.66	(0.85)			
SRS-22r Self- Image	0.59	0.53	(0.80)		
SRS-22r Mental Health	0.55	0.52	0.61	(0.87)	
SRS-22r Satisfaction	0.44	0.38	0.50	0.36	(0.70)
SRS-22r Total	0.82	0.81	0.83	0.81	0.60
SRS-20 Total	0.83	0.82	0.83	0.83	0.51
PROMIS Pain Intensity	−0.51	−0.80	−0.43	−0.34	−0.28
Mean	22.01	20.70	19.86	20.00	8.17
Standard Deviation	3.12	3.89	3.70	4.09	1.74
% Ceiling	28.0%	20.9%	11.3%	14.4%	32.8%

Note. Values on the diagonal in parentheses are Cronbach alpha estimates of internal consistency.

checked using jackknife slope index (JSI) [45]; item pairs with JSI more than 2 standard deviations away from the mean for each domain were flagged as being conditionally dependent. Plots of conditional standard errors of measure were examined to assess measurement precision for each domain; additionally, an overall empirical reliability of expected *a posteriori* scores was computed. Differential item functioning (DIF) was assessed using the lordif package for R [46]. DIF was assessed for sex; sample size was insufficient to test for DIF based on race.

Results

Data from 1823 patients who completed at least 1 SRS-22r questionnaire was used for psychometric analysis. The sample mean age was 14.9±2.2years, predominantly female (76.5%) and white, non-Hispanic (90%) race. The remaining racial distribution was 2% Hispanic, 5% Black/African American, and 3% other. Clinical parameters were not collected for the study, therefore makeup of the entire sample was not available, however, for the subset where data were available (N=840): data were from assessments completed predominately during an observation visit (79%), at a preoperative visit (10%) and at a postoperative visit (11%); curve patterns were 60% Thoracic, 23% Lumbar, and 17% Thoracolumbar; Risser grades were Risser 0 (17%), 1 (9%), 2 (8%), 3 (10%), 4 (32%), and 5 (24%); Curve magnitudes ranged from 5 degrees to 102 degrees (17% had curves >50degrees), with the mean curve 35±15 degrees. This distribution of patients is representative of the population of patients seen at the study facility and likely that of the entire dataset.

A table of inter-item correlations and item statistical properties are in Measurement Appendix A, and inter-subdomain correlations and subdomain statistical properties are in Table 1. Numerous items exhibit very strong ceiling effects. For items SRSq14 (self-image), SRSq15 (function), and SRSq17 (pain), over 80% of responses are in the top category. For items SRSq9 (function), SRSq11 (pain), SRSq12 (function), and SRSq18 (function), over 50% of responses are in the top category. Ceiling effects were found in several subdomains with a high percent of respondents responding in the maximum category for all items in that subdomain: function (28%), pain (21%), and satisfaction (33%). These ceiling effects are similar to findings reported by others [47–49].

The proposed structure of the SRS-22r, consisting of 5 related but distinct subdomains, was tested using ESEM. The ESEM demonstrated acceptable fit to the data: χ^2 (130)=343.73, $p < .001$; RMSEA=0.035; CFI=0.98; TLI=0.96; SRMR=0.02. No large (>0.10) residual correlations were found. The estimated ESEM parameters can be found in Table 2. A number of items failed to adequately load onto their assigned factor. In the function subdomain, items SRSq5, SRSq12, and SRSq15 had standardized factor loadings less than 0.40; of function items, only items SRSq9 and SRSq18 had adequate loadings. In the pain subdomain, item SRSq17 had a low loading; in the self-image subdomain, items SRSq4 and SRSq14 low loadings. In fact, the loading of SRSq14 onto the Self-Image factor was so small (0.045) as to be statistically nonsignificant. All other items loaded adequately onto their assigned factors.

Three crossloadings were found. Items SRSq14 (self-image) and SRSq17 (pain) crossload onto the function factor. SRSq12 (function) exhibited a crossloading onto the pain subdomain. Correlations in the ESEM were noticeably smaller than observed correlations in Table 2 between function and pain (0.66 vs. 0.51), function and self-image (0.59 vs. 0.32), and function and mental health (0.55 vs. 0.39).

Item and scale functioning were assessed by fitting graded response IRT models to each subdomain except Satisfaction, which only has 2 items. Item fit was found to be adequate for all items except SRSq10 (self-image), SRSq16 (mental health), and SRSq20 (mental health). Examination of the empirical item fit plots revealed only minor deviations, so the graded response model was deemed to adequately represent item responses. Two item pairs were found to exhibit local dependence: SRSq1 and SRSq2 (pain), and SRSq7 and SRSq16 (mental health). For each pair, a sensitivity analysis detailed in the Measurement Appendix A was performed to explore the severity of this local dependence. The local dependence between SRSq1 and SRSq2 was found to have a substantial effect on estimation of item parameters and score precision.

Item parameter estimates for the IRT models are in Table 3. Item discriminations are within normal levels for items in psychological measures, except items SRSq1 (pain), SRSq2 (pain), and SRSq16 (mental health). Notably, these items are all involved in a locally dependent item pair. Item thresholds are overall very low, reflecting the substantial ceiling effects of many items and moderate ceiling effects of the rest of the items. Five items (SRSq5, SRSq10, SRSq11, SRSq14, SRSq15) have a bottom threshold below −3.0, suggesting the bottom category is not well used.

Estimated trait scores and standard errors of estimate were computed for all subdomains (Figure 1). Empirical reliability of function was inadequate (0.69) and many participants throughout the score continuum were measured with substantial error. Empirical reliability of pain was high (0.89) and measurement precision is high for all participants except those responding in the top category to all items. Empirical reliability of self-Image was acceptable (0.84), and all participants were measured with acceptable to high levels of precision except for all participants except those responding in the top category to all items. For these participants, measurement precision was marginally acceptable. Empirical reliability of the mental health domain was high (0.87) and measurement precision is high for all participants except those responding in the top category to all items.

When items were assessed for variation between males and females, the only item flagged for DIF was SRSq8 (Pain; back pain when at rest); males have higher scores than females conditional upon level of the pain trait. However, the effect size of this difference was very small, dDTF=0.02 [50] and therefore of little concern.

Table 2
Standardized factor loadings from exploratory structural equation model of SRS-22r.

Domain	Question Number	SRS-22r Function	SRS-22r Pain	SRS-22r Self-Image	SRS-22r Mental Health	SRS-22r Satisfaction
Function	SRSq5	0.157	0.108	0.131	0.214	0.098
	SRSq9	**0.403**	0.077	0.118	0.055	0.032
	SRSq12	0.377	**0.421**	0.022	0.083	0.046
	SRSq15	0.282	−0.031	−0.062	0.155	0.075
	SRSq18	**0.487**	0.077	0.238	−0.076	0.111
Pain	SRSq1	0.019	**0.779**	0.076	0.064	−0.011
	SRSq2	0.033	**0.782**	0.054	0.03	0.016
	SRSq8	−0.155	**0.808**	−0.008	0.106	0.06
	SRSq11	0.265	**0.496**	0.024	−0.061	−0.066
	SRSq17	**0.429**	0.326	−0.121	−0.016	0.092
Self-Image	SRSq4	−0.081	0.254	0.376	−0.032	0.239
	SRSq6	0.051	−0.097	**0.848**	0.066	−0.071
	SRSq10	0.048	0.064	**0.761**	0.01	−0.018
	SRSq14	**0.347**	0.044	0.045	0.286	0.098
	SRSq19	0.057	−0.088	**0.648**	0.06	0.098
Mental Health	SRSq3	−0.065	0.068	0.114	**0.602**	−0.082
	SRSq7	−0.021	−0.012	−0.086	**0.888**	0.058
	SRSq13	0.022	0.157	0.156	**0.521**	−0.006
	SRSq16	0.028	−0.052	−0.073	**0.945**	−0.016
	SRSq20	0.044	−0.012	0.173	**0.568**	0.067
Satisfaction	SRSq21	−0.026	0.027	0.053	0.016	**0.761**
	SRSq22	0.056	−0.076	0.020	−0.018	**0.674**

Estimated Inter-factor Correlations

	Domain	Function	Pain	Self-Image	Mental Health
	Pain	0.506			
	Self-Image	0.322	0.472		
	Mental Health	0.386	0.518	0.593	
	Satisfaction	0.366	0.435	0.54	0.415

Discussion

Using modern psychometric analyses, this study generated additional robust evidence necessary to assist clinicians in understanding to what degree they can confidently interpret SRS-22r (English version) total and individual domain scores for adolescent patients with AIS. Addressing the primary study aim to improve interpretation and clinical utility of the SRS-22r for adolescent patients with scoliosis, the statistical results are now synthesized and the *clinical interpretations for translation into clinical and research practice* discussed. While the results were presented according to the study question and the type of validity evidence evaluated, the clinical interpretation of these findings are organized by content domain. Each SRS-22r domain is discussed separately and evaluated based on it appropriateness as a research tool and as a clinical tool.

SRS-22r function

Several items in the function subdomain were found to have high ceiling effects contributing to the concerningly low reliability of the function domain. SRSq15 (financial difficulties) has very low correlations with other function items but is not surprising considering adolescents cannot reasonably be expected to understand their household's financial situation. This item may therefore not be appropriate for adolescents. As a sensitivity analysis, all analyses reported herein were conducted with SRSq15 removed. These new analysis showed minimal differences from the original analyses; inferences from the ESEM and IRT model were identical and all correlation between the function domain and other variables were within 0.02 of the original estimate. The empirical reliability estimated from the IRT model was unchanged; however Cronbach's alpha estimate of internal consistency was slightly higher (0.71 compared to 0.69) after removing SRSq15. The minimal changes upon removal of SRSq15 are not surprising given the extreme ceiling effect (92%) renders SRSq15 essentially constant.

The function domain may not be a pure representation of quality of physical functioning with scoliosis due to the numerous crossloadings found involving the function factor. Similar crossloadings have been found in factor analytic studies of the Japanese [16], Korean [51], French [52], French-Canadian [19] and Spanish [53] versions of the SRS-22r. Statistically, correlations between SRS-22r domains were significantly and substantially different when computed with total scores compared to inter-factor correlations within the ESEM model. Accordingly, researchers should be aware that correlations using observed subdomain scores may not accurately reflect correlations between the subdomain constructs, particularly for the function subdomain.

Clinical uses of function. Given that ceiling effects were significant for 4 of the 5 function items, individuals without problems with physical functioning are likely to have perfect or nearly perfect function scores. Since typically AIS does not significantly affect one's functional ability and has been seen as a cosmetic condition, the ceiling effects for function items are neither unexpected nor necessarily problematic from a clinical perspective. Specifically, since the clinical use of these scores is often to identify patients with atypical physical function relative to AIS peers, using the domain score is appropriate to guide the physician to look further into explaining the patient's difficulty with physical function. Once a patient is identified as having atypical physical function, low scores on individual function items may be a reasonable tool to inform the clinician of a physical functioning problem to direct treatment interventions if necessary. However, given the limitation of the function domain score, clinicians specifically concerned about physical functioning should consider utilizing a supplemental instrument focusing on sport or higher physical functioning.

Research uses of function. Low reliability and possibly biased correlations with other SRS-22r domains make the function domain ill-suited for use in correlational analyses (e.g., correlations, regression, path analysis, etc). Because this study did not include a longitudinal component, we are unable to comment on how these issues do or do not affect the function domain's sensitivity to change over time. A study focusing on

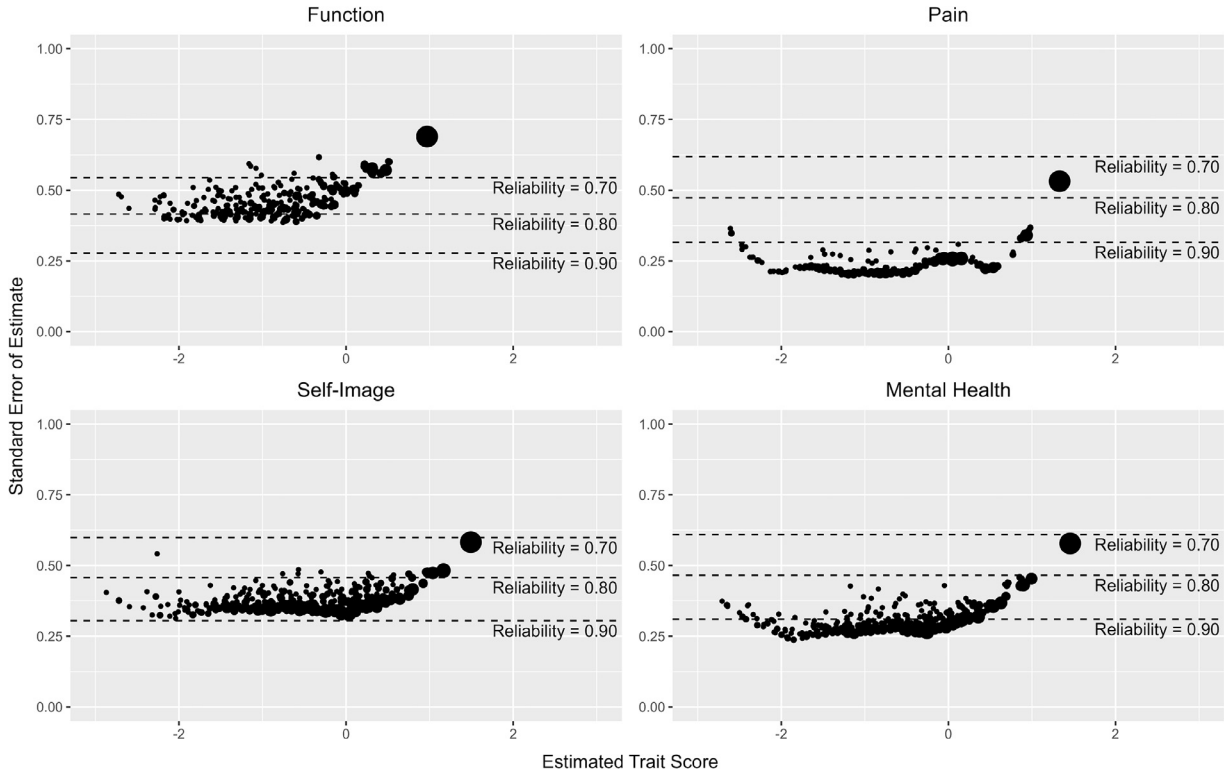

Figure 1. Conditional Standard Errors of Estimate for SRS-22r subdomains.

variability within persons and across time would be better suited to addressing the stability of function and its sensitivity to change. However, given the low reliability of the function domain, it is to be expected that power to detect changes over time (e.g., treatment effects) would be low, and estimating effect sizes for those changes would be difficult [54].

We propose that the function domain could be improved to be more representative of physical function by removing items SRSq15 and SRSq18 (and replacing them with new items intended to measure level of physical function, e.g., ability to play sports, ability to lift heavy items, or ability to run and jump). This change would remove poorly performing items and help clarify the meaning of the function domain.

SRS-22r pain

High ceiling effects were found for 2 items (SRSq11, SRSq17) in the pain subdomain and as a result, these items do not strongly contribute to measurement of the pain subdomain, further evidenced by their lower discriminations in the item response theory analysis. Compounding this issue, SRSq1 (pain last 6 months) and SRSq2 (pain last month) are more strongly related to each other than to the other questions in the domain. As a result, they strongly influence the pain domain score, and therefore are the primary determinants of any interpretation of the pain domain score. Indeed, the very high correlation between the pain subdomain and PROMIS pain intensity indicates the pain domain measures pain intensity rather than the extent to which back pain influences quality of life with scoliosis. That said the pain subdomain has excellent internal consistency and empirical reliability.

Clinical uses of pain. Given that ceiling effects were significant for 2 of the 5 pain items, individuals without pain problems are likely to have very high pain domain scores, relflecting this lack of pain. Similar to the function domain, the ceiling effects for pain items are not unexpected or necessarily problematic, since typically, patients with AIS have not reported pain as a leading problem. Specifically, since the clinical use of these scores is often to identify patients with atypical levels of pain relative to AIS peers using the domain score is appropriate to guide the physician to look further into explaining the patient's pain. Once a patient is identified as having atypical pain, low scores on individual pain items may be a reasonable diagnostic tool to inform the clinician of the nature pain problem. Compared to the PROMIS pain intensity item, the SRS-22r pain domain items involve a longer time frame, which can help clinicians categorize patients' pain as chronic vs. acute. A related study [55] discusses in more detail the specifics of pain scores across different patient reported outcomes in patients with AIS.

Research uses of pain. Psychometric data suggests that the pain domain scores should be an effective research tool for assessing pain *intensity*. The high reliability suggests pain scores will have good sensitivity to change and be useful for correlation research. Future work should investigate the performance of the pain domain in longitudinal settings and in settings involving treatment.

SRS-22r self-image

The interpretation of SRS-22r self-image scores should be viewed as a measure of how the individual feels about their appearance. All items ask the patient about appearance except SRSq14 which demonstrates a weak connection to the self-image domain in the SEM and IRT models. The weak correlation for item SRSq4 is likely due to the question structure that has a lack of emphasis on the word "shape." The remainder of the 25-word item reads as a general quality of life item; the authors suspect it is easy for an adolescent to overlook the word "shape," thereby misinterpreting the question and changing its meaning entirely.

Clinical uses of self-image. Clinicians should be aware of the emphasis of appearance when interpreting the self-image score clinically. It is recommended to focus on individual item responses if this domain yields a concerningly low score for a patient. Overall, the self-image domain score is useful to inform clinicians on how the patient perceives their scoliosis and its influence on their appearance and can be useful for planning future treatment or evaluating treatment choices such as bracing. It is important to note that in other work, self-image scores have

Table 3
Item response theory parameter estimates for SRS-22r domains.

SRS-22r Function					
Item	Discrimination	Thresholds			
		b1	b2	b3	b4
SRS5	1.42	−4.98	−2.31	−0.81	0.06
SRS9	1.75	−2.85	−2.36	−1.72	−0.82
SRS12	2.65	−2.48	−1.87	−0.98	−0.28
SRS15	1.11	−5.79	−4.16	−3.55	−0.27
SRS18	1.58	−2.83	−2.08	−0.85	−0.41
SRS-22r Pain					
Item	Discrimination	Thresholds			
		b1	b2	b3	b4
SRS1	7.78	−2.08	−1.17	−0.44	0.60
SRS2	6.82	−2.05	−1.26	−0.58	0.42
SRS8	2.27	−1.93	−1.17	−0.30	0.51
SRS11	1.79	−3.88	−2.80	−2.21	−0.67
SRS17	1.92	−2.49	−2.26	−1.89	−1.47
SRS-22r Self-Image					
Item	Discrimination	Thresholds			
		b1	b2	b3	b4
SRS4	1.39	−2.02	−1.08	0.13	1.05
SRS6	3.21	−2.65	−2.01	−0.67	0.33
SRS10	2.99	−3.01	−1.91	−0.49	0.62
SRS14	1.48	−4.81	−3.37	−2.41	−1.38
SRS19	2.34	−2.07	−1.49	−0.48	0.47
SRS-22r Mental Health					
Item	Discrimination	Thresholds			
		b1	b2	b3	b4
SRS3	1.84	−2.44	−1.51	−0.62	0.54
SRS7	3.47	−2.13	−1.65	−0.83	0.01
SRS13	2.31	−2.61	−1.67	−0.63	0.83
SRS16	4.55	−2.10	−1.61	−0.74	0.03
SRS20	2.43	−2.86	−1.81	−0.94	0.51

not been strongly correlated with curve magnitude [56]; therefore, self-image scores should be considered as an assessment of the patient's perception of their appearance rather than an objective measure of curve severity. Despite the limitations, this self-image domain remains a useful domain in clinical management.

Research uses of self-image. Results suggest SRS-22r self-image scores should be interpreted narrowly, as the patient's perception of their appearance, rather than more generally as the patient's self-image. A sensitivity analysis was performed by removing SRSq4 and SRSq14 (which do not address appearance) from the self-image scores to determine whether any results would be different. None of the analyses yielded meaningfully different results: correlations amongst subdomains were unchanged, parameters in ESEM and IRT models were largely unchanged, and estimates of reliability were nearly identical. For research focusing on self-image or appearance, we recommend supplementing the SRS-22r self-image with other research tools such as the Spinal Appearance Questionnaire (SAQ) or the Harters Self-Perception Profile. Further research focusing on the interpretation and validity of SRS-22r self-image scores, including its sensitivity to change, are recommended.

Mental health

The mental health domain has excellent psychometric properties such as high reliability and little ceiling effect. Furthermore, based on moderate interdomain correlations and lack of crossloadings in the ESEM, mental health scores are distinct from scores of other domains. Mental health is an important domain to assess in this population, as evidenced by increased reports of depression and anxiety [57]. The use and interpretation of the domain scores in both the clinical and research contexts is supported by this work.

SRS-22r total score

SRS-22r total score, is a composite score comprised of averaging the 4 individual domain scores, as a representation a singular quality of life score. Based on the only moderate inter-factor correlations found in the ESEM, the authors caution about interpreting an SRS-22r total score in this manner. Additionally, more detailed analyses conducted and reported in the Measurement Appendix A suggest that each subdomain captures significant unique variance, making a unidimensional interpretation of SRS-22r data problematic. Prior researchers similarly found too much heterogeneity across SRS-22r items to admit a unidimensional interpretation and suggested a shortened, 7-item, scale which they found to be adequately unidimensional [31,32,58]. Our results, along with the qualitative study by Alamrani et al. [6,33], that found the SRS-22r does not adequately capture important concepts that relate to HRQoL of adolescents with AIS, suggest a single score as a proxy for this construct may not be ideal.

Clinical uses of total score. Clinically, the value of SRS-22r scores comes from comparing individual domain scores to prior scores from the same individual and to published domain norms. The individual domain scores contain unique information that may be lost by combining them into a single total score. It is, however, possible that an overall score would have a greater responsiveness to change than any of the individual domain scores, but that issue is beyond the scope of this project.

Research use of total score. Creation of a singular score that encompasses the construct of overall quality of life is a difficult, especially given the many different definitions and measurement strategies [59]. Therefore, users of any quality of life measure should be clear about their own definition and empirical conceptualization of quality of life. For researchers conceptualizing health-related quality of life in a formative manner, such that a patient's experience of pain, function, self-image, and mental health are seen as causing that patient's perception of overall quality of life, then quality of life should be treated as a multidimensional composite variable. Specifically, a multivariate analytical approach such as multivariate analysis of (co-)variance (MANOVA/MANCOVA) could be used to assess differences and/or changes in overall quality of life. Caution should be employed, however, as this study (nor any other that we are aware of) does not consider the validity of the multidimensional composite approach to measuring health-related quality of life.

Study strengths and limitations

Traditionally, implications of advanced psychometric analyses can be difficult to translate into clinical practice. This paper has not only provided the critical statistical results, but also interpreted the results of statistical analyses, considering equally the perspectives the clinician and the researcher, understanding the uniqueness in the application of the use of the SRS-22r. This approach should improve the interpretation and clinical utility of the SRS-22r for adolescent patients with AIS.

The analyses for this study included a sample that was (1) large, (2) restricted to patient with an AIS diagnosis, (3) heterogeneous (sample size was designed to capture enough males to enable comparison of measurement properties across sex), and (4) avoided possible within-persons contamination by randomly selecting a single measurement occasion for each person. We are unaware of any previous psychometric analyses of the SRS-22r with as strong a patient sample.

The statistical analyses employed for this study are the most in depth and modern psychometric analyses of the SRS-22r (English) in recent years. The use of ESEM enables a detailed analysis of common variance of items, enabling us to determine whether each item was properly measuring the domain to which it belongs. This technique is similar to

exploratory factor analysis (EFA) which has been previously used; however, since the factor structure is prespecified, it provides a more robust test of the theoretical partitioning of items into factors. More specifically, ESEM can be seen as a hypothesis testing approach to factor structure, whereas EFA is purely hypothesis generating. Furthermore, the use of item response theory enabled an assessment of measurement quality across the continuum of scores (i.e., rather than a single overall reliability estimate); furthermore, the use of discriminations from IRT allowed us to identify items with overmuch (or not enough) influence on their domain.

This study is not without limitations. Many aspects of validity were beyond the scope of this study, as it was cross-sectional and did not include variables beyond the SRS-22r and patient demographics. As such, content validity, responsiveness, test-retest reliability, and other externally focused sources of validity evidence were not considered. The authors feel this limitation is warranted because our overarching goal was to aid researchers in determining how best to use the SRS-22r instrument and interpret scores arising from that usage. By focusing solely on the SRS-22r items, we were able to provide valuable insight into the functioning and possible of interpretations of SRS-22r subdomains. These insights, in turn, will inform any interpretations of statistical tests using the SRS-22r.

Conclusions

Modern psychometric assessment of the SRS-22r, in a large group of pediatric patients with AIS, were presented and interpreted to assist clinicians and researchers in understanding the strengths and limitations of this commonly used PRO. Overall, the SRS-22r demonstrated good psychometric properties in all domains except function. This work draws attention to shortcomings of the SRS-22r that future researchers can address through modification similar to the conversion from SRS-24 to SRS-22r (2003). These modifications should support conversion between different versions as in the equating study (2006) accompanying the release of the SRS-22r. Given the SRS-22r is the primary scoliosis HRQoL PRO, it is essential to ensure it is of the highest quality while simultaneously working well with archival data so that longitudinal studies can inform treatment and care. In its current form, it is critical that clinicians and researchers be informed of its strengths and limitations, and the implications for management of their patients with AIS.

Declarations of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This study was funded by Kosair for Kids (formerly Kosair Charities, Inc.).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.xnsj.2024.100545.

References

[1] Goldberg MS, Mayo NE, Poitras B, Scott S, Hanley J. The Ste-Justine adolescent idiopathic scoliosis cohort study. Part I: description of the study. Spine (Phila Pa 1976) 1994;19(14):1551–61.
[2] Alonso J, Bartlett SJ, Rose M, et al. The case for an international patient-reported outcomes measurement information system (PROMIS[R]) initiative. Health Qual Life Outcomes 2013;11:210.
[3] Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract 2006;12(5):559–68.
[4] Harshavardhana NS, Dormans J. Review of patient reported outcome measures used in pediatric spinal deformity surgery. University of Pennsylvania Orthopaed J 2013;23(June):31–2.
[5] Quittner AL, Nicolais CJ, Saez-Flores E, et al. 13 - integrating patient-reported outcomes into research and clinical practice. In: Wilmott RW, Deterding R, Li A, et al., editors. Kendig's disorders of the respiratory tract in children. Philadelphia: Elsevier; 2019. p. 231–40. .e3.
[6] Alamrani S, Gardner A, Falla D, Russell E, Rushton AB, Heneghan NR. Content validity of Scoliosis Research Society questionnaire-22 revised (SRS-22r) for adolescents with idiopathic scoliosis: protocol for a qualitative study exploring patient's and practitioner's perspectives. BMJ Open 2021;11(12):e053911.
[7] Haher TR, Gorup JM, Shin TM, et al. Results of the Scoliosis Research Society instrument for evaluation of surgical outcome in adolescent idiopathic scoliosis. A multicenter study of 244 patients. Spine (Phila Pa 1976) 1999;24(14):1435–40.
[8] Asher M, Min Lai S, Burton D, Manna B. Scoliosis research society-22 patient questionnaire: responsiveness to change associated with surgical treatment. Spine (Phila Pa 1976) 2003;28(1):70–3.
[9] Asher M, Min Lai S, Burton D, Manna B. The reliability and concurrent validity of the scoliosis research society-22 patient questionnaire for idiopathic scoliosis. Spine (Phila Pa 1976) 2003;28(1):63–9.
[10] Alamrani S, Rushton AB, Gardner A, Bini E, Falla D, Heneghan NR. Physical functioning in adolescents with idiopathic scoliosis: a systematic review of outcome measures and their measurement properties. Spine (Phila Pa 1976) 2021;46(18):E985–EE97.
[11] Society SR. Patient outcome questionnaires. [cited 2023 7/18/2023]; Available from: https://www.srs.org/professionals/online-education-and-resources/patient-outcome-questionnaires.
[12] Bago J, Climent JM, Ey A, Perez-Grueso FJ, Izquierdo E. The Spanish version of the SRS-22 patient questionnaire for idiopathic scoliosis: transcultural adaptation and reliability analysis. Spine (Phila Pa 1976) 2004;29(15):1676–80.
[13] Monticone M, Carabalona R, Negrini S. Reliability of the Scoliosis Research Society-22 Patient Questionnaire (Italian version) in mild adolescent vertebral deformities. Eura Medicophys 2004;40(3):191–7.
[14] Alanay A, Cil A, Berk H, et al. Reliability and validity of adapted Turkish Version of Scoliosis Research Society-22 (SRS-22) questionnaire. Spine (Phila Pa 1976) 2005;30(21):2464–8.
[15] Cheung KM, Senkoylu A, Alanay A, Genc Y, Lau S, Luk KD. Reliability and concurrent validity of the adapted Chinese version of Scoliosis Research Society-22 (SRS-22) questionnaire. Spine (Phila Pa 1976) 2007;32(10):1141–5.
[16] Hashimoto H, Sase T, Arai Y, Maruyama T, Isobe K, Shouno Y. Validation of a Japanese version of the Scoliosis Research Society-22 patient questionnaire among idiopathic scoliosis patients in Japan. Spine (Phila Pa 1976) 2007;32(4):E141–6.
[17] Zhao L, Zhang Y, Sun X, Du Q, Shang L. The Scoliosis Research Society-22 questionnaire adapted for adolescent idiopathic scoliosis patients in China: reliability and validity analysis. J Child Orthop 2007;1(6):351–5.
[18] Antonarakos PD, Katranitsa L, Angelis L, et al. Reliability and validity of the adapted Greek version of Scoliosis Research Society - 22 (SRS-22) questionnaire. Scoliosis 2009;4:14.
[19] Beausejour M, Joncas J, Goulet L, et al. Reliability and validity of adapted French Canadian version of Scoliosis Research Society Outcomes Questionnaire (SRS-22) in Quebec. Spine (Phila Pa 1976) 2009;34(6):623–8.
[20] Glowacki M, Misterska E, Laurentowska M, Mankowski P. Polish adaptation of Scoliosis Research Society-22 questionnaire. Spine (Phila Pa 1976) 2009;34(10):1060–5.
[21] Niemeyer T, Schubert C, Halm HF, Herberts T, Leichtle C, Gesicki M. Validity and reliability of an adapted German version of Scoliosis Research Society-22 questionnaire. Spine (Phila Pa 1976) 2009;34(8):818–21.
[22] Mousavi SJ, Mobini B, Mehdian H, et al. Reliability and validity of the Persian version of the Scoliosis Research Society-22r questionnaire. Spine (Phila Pa 1976) 2010;35(7):784–9.
[23] Rosanova GC, Gabriel BS, Camarini PM, Gianini PE, Coelho DM, Oliveira AS. Concurrent validity of the Brazilian version of SRS-22r with Br-SF-36. Rev Bras Fisioter 2010;14(2):121–6.
[24] Lee JS, Lee DH, Suh KT, Kim JI, Lim JM, Goh TS. Validation of the Korean version of the Scoliosis Research Society-22 questionnaire. Eur Spine J 2011;20(10):1751–6.
[25] Leelapattana P, Keorochana G, Johnson J, Wajanavisit W, Laohacharoensombat W. Reliability and validity of an adapted Thai version of the Scoliosis Research Society-22 questionnaire. J Child Orthop 2011;5(1):35–40.
[26] 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(21):1875–84.
[27] Schlosser TP, Stadhouder A, Schimmel JJ, Lehr AM, van der Heijden GJ, Castelein RM. Reliability and validity of the adapted Dutch version of the revised Scoliosis Research Society 22-item questionnaire. Spine J 2014;14(8):1663–72.
[28] Haidar RK, Kassak K, Masrouha K, Ibrahim K, Mhaidli H. Reliability and validity of an adapted Arabic version of the Scoliosis Research Society-22r Questionnaire. Spine (Phila Pa 1976) 2015;40(17):E971–7.
[29] Asher MA, Lai SM, Glattes RC, Burton DC, Alanay A, Bago J. Refinement of the SRS-22 Health-Related Quality of Life questionnaire function domain. Spine (Phila Pa 1976) 2006;31(5):593–7.
[30] Glattes RC, Burton DC, Lai SM, Frasier E, Asher MA. The reliability and concurrent validity of the Scoliosis Research Society-22r patient questionnaire compared with the Child Health Questionnaire-CF87 patient questionnaire for adolescent spinal deformity. Spine (Phila Pa 1976) 2007;32(16):1778–84.
[31] Caronni A, Zaina F, Negrini S. Improving the measurement of health-related quality of life in adolescent with idiopathic scoliosis: the SRS-7, a Rasch-developed short form of the SRS-22 questionnaire. Res Dev Disabil 2014;35(4):784–799.

[32] Jain A, Sponseller PD, Negrini S, et al. SRS-7: a valid, responsive, linear, and unidimensional functional outcome measure for operatively treated patients with AIS. Spine (Phila Pa 1976) 2015;40(9):650–5.
[33] Alamrani S, Gardner A, Falla D, Russell E, Rushton AB, Heneghan NR. Content validity of the Scoliosis Research Society questionnaire (SRS-22r): a qualitative concept elicitation study. PLoS One 2023;18(5):e0285538.
[34] Marsh HW, Morin AJ, Parker PD, Kaur G. Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. Annu Rev Clin Psychol 2014;10:85–110.
[35] Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. Patient 2014;7(1):23–35.
[36] Stover AM, McLeod LD, Langer MM, Chen WH, Reeve BB. State of the psychometric methods: patient-reported outcome measure development and refinement using item response theory. J Patient Rep Outcomes 2019;3(1):50.
[37] Muthén L, Muthén B. Mplus user's guide.
[38] Asparouhov T, Muthén B. Exploratory structural equation modeling. Struct Equat Model: A Multidiscipl J 2009;16(3):397–438.
[39] Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociol Meth Res 2016;21(2):230–58.
[40] Sharma S, Mukherjee S, Kumar A, Dillon WR. A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. J Bus Res 2005;58(7):935–43.
[41] Byrne BM. Structural equation modeling with LISREL, PRELIS, and SIMPLIS: basic concepts, applications, and programming. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 1998.
[42] Kline RB. Principles and practice of structural equation modeling, New York, NY: The Guilford Press; 2016. ed. t, ed..
[43] Howard MC. A review of exploratory factor analysis decisions and overview of current practices: what we are doing and how can we improve? Int J Human-Computer Interact 2015;32(1):51–62.
[44] Orlando M, Thissen D. Likelihood-based item-fit indices for dichotomous item response theory models. Appl Psychol Measurem 2016;24(1):50–64.
[45] Edwards MC, Houts CR, Cai L. A diagnostic procedure to detect departures from local independence in item response theory models. Psychol Methods 2018;23(1):138–49.
[46] Choi SW, Gibbons LE, Crane PK. lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and monte carlo simulations. J Stat Softw 2011;39(8):1–30.
[47] Lai SM, Asher MA, Burton DC, Carlson BB. Identification of Scoliosis Research Society-22r health-related quality of life questionnaire domains using factor analysis methodology. Spine (Phila Pa 1976) 2010;35(12):1236–40.
[48] Mitchell SL, McLaughlin KH, Bachmann KR, Sponseller PD, Reider LM. Construct validity of pediatric PROMIS computerized adaptive testing measures in children with adolescent idiopathic scoliosis. J Pediatr Orthop 2022;42(7):e720–e7e6.
[49] Stone LE, Ames CP, Pellise F, et al. Scoliosis Research Society-22r and ceiling effects: limited capabilities for precision-medicine with adolescent idiopathic scoliosis. Spine (Phila Pa 1976) 2024. doi:10.1097/BRS.0000000000004993.
[50] Stark S, Chernyshenko OS, Drasgow F. Examining the effects of differential item (functioning and differential) test functioning on selection decisions: when are statistically significant effects practically important? J Appl Psychol 2004;89(3):497–508.
[51] Lee H, Choi J, Hwang JH, Park JH. Psychometric evaluation of the Scoliosis Research Society-22 Revised questionnaire among adolescents with idiopathic scoliosis. J Pediatr Orthop B 2017;26(1):59–63.
[52] Theroux J, Stomski N, Innes S, et al. Revisiting the psychometric properties of the Scoliosis Research Society-22 (SRS-22) French version. Scoliosis Spinal Disord 2017;12:21.
[53] Climent JM, Bago J, Ey A, Perez-Grueso FJ, Izquierdo E. Validity of the Spanish version of the Scoliosis Research Society-22 (SRS-22) patient questionnaire. Spine (Phila Pa 1976) 2005;30(6):705–9.
[54] Cole DA, Preacher KJ. Manifest variable path analysis: potentially serious and misleading consequences due to uncorrected measurement error. Psychol Methods 2014;19(2):300–15.
[55] Griffin J, Bagley A, Iwinski H, et al. Translating patient-reported outcomes (PROs) into clinical practice: an investigation of pain-related pros in adolescent idiopathic scoliosis.
[56] Bouton D, Fedorak G, Oeffinger DJ, et al. Baseline patient reported outcome measurement information system (PROMIS) scores in children with idiopathic scoliosis and their relation to the SRS-22. Spine Deform 2022;10(1):63–8.
[57] Mitsiaki I, Thirios A, Panagouli E, et al. Adolescent idiopathic scoliosis and mental health disorders: a narrative review of the literature. Children (Basel) 2022;9(5):597. doi:10.3390/children9050597.
[58] Lai SM, Burton DC, Asher MA, Carlson BB. Converting SRS-24, SRS-23, and SRS-22 to SRS-22r: establishing conversion equations using regression modeling. Spine (Phila Pa 1976) 2011;36(23):E1525–33.
[59] Lin X-J, Lin IM, Fan S-Y. Methodological issues in measuring health-related quality of life. Tzu Chi Med J 2013;25(1):8–12.