Eur Spine J (2018) 27:685–699
https://doi.org/10.1007/s00586-017-5279-0

ORIGINAL ARTICLE

Factor analysis of the SRS-22 outcome assessment instrument in patients with adult spinal deformity

A. F. Mannion[1] · A. Elfering[2] · J. Bago[3] · F. Pellise[3] · A. Vila-Casademunt[4] · S. Richner-Wunderlin[1] · M. Domingo-Sàbat[4] · I. Obeid[5] · E. Acaroglu[6] · A. Alanay[7] · F. S. Pérez-Grueso[8] · C. R. Baldus[9] · L. Y. Carreon[10] · K. H. Bridwell[9] · S. D. Glassman[10] · F. Kleinstück[11] · European Spine Study Group (ESSG)[4]

Received: 28 March 2017 / Revised: 3 August 2017 / Accepted: 19 August 2017 / Published online: 2 September 2017

Abstract

Purpose Designed for patients with adolescent idiopathic scoliosis, the SRS-22 is now widely used as an outcome instrument in patients with adult spinal deformity (ASD). No studies have confirmed the four-factor structure (pain, function, self-image, mental health) of the SRS-22 in ASD and under different contexts. Factorial invariance of an instrument over time and in different languages is essential to allow for precise interpretations of treatment success and comparisons across studies. This study sought to evaluate the invariance of the SRS-22 structure across different languages and sub-groups of ASD patients.

Methods Confirmatory factor analysis was performed on the 20 non-management items of the SRS-22 with data from 245 American English-, 428 Spanish-, 229 Turkish-, 95 French-, and 195 German-speaking patients. Item loading invariance was compared across languages, age groups, etiologies, treatment groups, and assessment times. A separate sample of SRS-22 data from 772 American surgical patients with ASD was used for cross-validation.

Results The factor structure fitted significantly better to the proposed four-factor solution than to a unifactorial solution. However, items 14 (personal relationships), 15 (financial difficulties), and 17 (days off work) consistently showed weak item loading within their factors across all language versions and in both baseline and follow-up datasets. A trimmed SRS (16 non-management items) that used the four least problematic items in each of the four domains yielded better-fitting models across all languages, but equivalence was still not reached. With this shorter version there was equivalence of item loading with respect to treatment (surgery vs conservative), time of assessment (baseline vs 12 months follow-up), and etiology (degenerative vs idiopathic), but not age ($<$ vs $\geq$50 years). All findings were confirmed in the cross-validation sample.

Conclusion We recommend removal of the worst-fitting items from each of the four domains of the SRS-instrument (items 3, 14, 15, 17), together with adaptation and

✉ A. F. Mannion
anne.mannion@yahoo.com

[1] Spine Center Division, Department of Teaching, Research and Development, Schulthess Klinik, Lengghalde 2, 8008 Zurich, Switzerland

[2] Institute for Psychology, University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland

[3] Spine Unit, Hospital Universitari Vall Hebron, Passeig Vall Hebron 119-129, Traumatology Building 2nd Floor, 08035 Barcelona, Spain

[4] Spine Research Unit, Vall Hebron Institute of Research (VHIR), Passeig Vall Hebron 119-129, Traumatology Building 2nd Floor, 08035 Barcelona, Spain

[5] Pellegrin Bordeaux University Hospital, Place Amélie Raba Léon, 33000 Bordeaux, France

[6] Ankara Spine Center, Iran Caddesi 45/2, Kavaklidere, 06700 Ankara, Turkey

[7] Department of Orthopaedics and Traumatology, Acibadem University School of Medicine, Büyükdere cad, 40 Maslak, 344457 Istanbul, Turkey

[8] Hospital Universitario La, Paz Paseo de la Castellana 261, 28046 Madrid, Spain

[9] Department of Orthopedics, Washington University School of Medicine, St. Louis, MO 63110, USA

[10] Norton Leatherman Spine Center, 210 East Gray Street, Suite 900, Louisville, KY 40205, USA

[11] Spine Center, Schulthess Klinik, Lengghalde 2, 8008 Zürich, Switzerland

standardization of other items across language versions, to provide an improved version of the instrument with just 16 non-management items.

Keywords Adult deformity · Scoliosis Research Society-22 (SRS-22) · Outcome · Factor analysis · Structural and cross-cultural validity

Introduction

It is now generally accepted that patients' perceptions of their health-related quality of life (HRQL) are of utmost importance when assessing the outcome of treatment for spinal disorders. To this end, various condition-specific instruments have been developed. One such instrument, the Scoliosis Research Society-22 (SRS-22) questionnaire, was designed for assessing health-related quality of life in patients with adolescent idiopathic scoliosis (AIS) [1–5]. The questionnaire has subsequently become the outcome instrument of choice also in patients with adult spinal deformity (ASD). The SRS-22 is a multidimensional instrument covering the four non-management domains of pain, function, self-image, and mental health (five items per domain), along with two items to assess the patient's satisfaction with management of their condition. The domains selected for inclusion were believed to be relevant in capturing the wide-ranging impact of scoliosis on various aspects of the patient's life. On the back of shortcomings in the factor structure of the instrument identified by Rasch analyses of data from AIS patients [6–8], a shorter, unidimensional, and linearly scaled 7-item version (SRS-7) of the instrument was recently proposed [9]. However, the authors conceded that although this was useful for assessing global changes in patient outcomes over time, a multidimensional instrument was likely better for assessing changes associated with individual aspects of the disease, such as cosmesis, pain, and function [9]. The longer version therefore continues to be considered an appropriate outcome instrument. Since its introduction, no studies have confirmed the four-factor structure of the non-management items of the SRS-22 in patients with ASD, or evaluated the validity of this structure in different contexts (different aetiologies of deformity, age-groups, languages, time-points of assessment, etc.). There are still many unknowns in the diagnosis, classification and treatment of adult spinal deformity, and progress in the evaluation and management of these disorders requires the pooling and sharing of data in large multicentre, international studies and registries. Since spinal deformities are often long-lasting conditions, it is also necessary that longitudinal studies of change are carried out over extended periods of time, to monitor the natural history and/or response to treatment. Demonstration of factorial invariance of the SRS-22 instrument in different languages, diagnostic sub-groups and age-groups, and also over time, is therefore essential if we are to make precise interpretations of treatment success and draw comparisons across studies.

Qualitative examination of the various official language versions of the SRS-22 reveals subtle differences in the wording used for some of the items and challenges the content validity of others. The differences appear to have arisen during the various refinements made to the SRS-22 over time (but not necessarily made concurrently in all current language versions). It is not known whether this has an impact on the psychometric properties of the instrument.

A recent systematic review of the psychometric properties of 17 published translations ("official versions") of the SRS-22 instrument [10] concluded that further attention should be given to the construct validity of the SRS-22 in terms of its cross-cultural validity (i.e., the extent to which the performance of an item in an adapted instrument adequately reflects its performance in the original version) and structural validity (i.e., the extent to which the scores adequately reflect the dimensional nature of the construct being measured).

The aim of this study was to evaluate the factor structure of the English [2, 5], Spanish [11, 12], French [13], Turkish [14], and German [15] versions of the SRS-22 outcome instrument in patients with ASD. We aimed to evaluate whether these versions showed the same four-factor structure for the non-management domains as the original version of the instrument created for AIS patients and whether this structure was invariant over time, etiology of deformity, treatment group, and patient-age.

Methods

Analyses were carried out using the 20 non-management items of the SRS-22[1] from baseline and 12-month questionnaire data. Data in the Spanish, Turkish, French and German languages came from a total of 947 patients with adult spinal deformity (recruited into the European Spine Study Group's (ESSG) prospective multicentre study) (Table 1). The ESSG comprises representatives from seven spine centres in Spain, Turkey, France, and Switzerland that collectively administer a multicentre international database of data from patients over 18 years of age with

[1] The official Spanish and Turkish versions contained the SRS-22R formulation of item 18, whilst that of SRS-22 was used in all other languages (http://www.srs.org/professionals/online-education-and-resources/patient-outcome-questionnaires).

spinal deformity defined as any one of the following: coronal plane deformity >20°; sagittal vertical axis >50 mm; pelvic tilt >25°; thoracic kyphosis >60° [16]. A comparable data set containing SRS-22 data in the English language ($N = 245$) from the "Multicenter, Prospective Adult Symptomatic Lumbar Scoliosis (ASLS)" study, funded by the National Institutes for Health (R01AR055176-01A2) in the USA (Table 1), was included to allow comparison with data collected in the original language version of the SRS questionnaire.

A third English language data set ($N = 772$; 645 (84%) female, age 50 ± 17 years), from the USA Adult Deformity Outcomes (ADO) database, was used for subsequent cross-validation of the models.

Statistical analyses

The factor structure of the 20 non-management items of the SRS-22 questionnaire was tested using confirmatory factor analysis (CFA), with structural equation modelling software AMOS 24.0. CFA assesses the contribution of each of the questionnaire's questions or "items" ("Item Loading") and measures the adequacy of the measurement model ("Goodness of Fit"). Item loading indicates the strength of the relationship between each item and its underlying factor, where values <0.6 are considered poor (analogous to a correlation coefficient). The various Goodness of Fit measures included the root mean square error of approximation (RMSEA), the ratio of Chi squared to degrees of freedom (χ^2/df) and the Comparative Fit Index (CFI). A model is considered to have a good fit if RMSEA is less than 0.05, χ^2/df is less than 2, and CFI is greater than 0.9 [17].

The analyses comprised systematic comparisons of different models. The comparability of the language versions was tested by constraining (forcing) item loading to be equal across languages and then testing whether this constrained model was as good a fit to the empirical data as one in which items were allowed to load "freely" (unconstrained) and potentially differ between the languages. The Chi squared difference test was used to assess the significance of the difference between constrained and unconstrained models; no significant difference confirmed equivalence of item loading for the different languages. The same type of model comparisons (i.e. item loading "constrained to be the same" in different subgroups versus "free, unconstrained") were made for the two measurement time-points (baseline and 12 months' follow-up) and for various clinical sub-groups (degenerative and idiopathic etiologies; ≤50 years and >50 years old; conservative and surgical patients). Again, Chi squared difference tests were used to assess equivalence of item loading across these groups.

The following models were evaluated:

Model 1 was a one-factor model that assumed all items loaded on a common factor i.e. "all 20 items measured the same construct".

Model 2 was the hypothesized four-factor structure with five questions per factor and with item loading estimated for the total sample; in model 2a item loading was estimated freely for the Spanish ($n = 428$), Turkish ($n = 229$), French ($n = 95$), German ($n = 195$), and English ($n = 245$) subsamples separately.

Model 3 comprised the best fitting item loading that represented all five language versions, i.e. items were constrained to have the same loading across the different languages.

After analysing these three models, four items were identified as being weak, i.e., they consistently had a low loading on their factor in all five languages and/or in baseline and follow-up data (see later). As such, an item-trimmed model, with just four items per factor was proposed (from hereon in, the 16-item version). Omission of the weakest items from model 2 was expected to significantly improve the model fit.

Model 4 comprised the four-factor structure for the 16-item version, with the item loading estimated for the total sample; in model 4a, item loading was estimated freely for the Spanish ($n = 428$), Turkish ($n = 229$), French ($n = 95$), German ($n = 195$), and English ($n = 245$) subsamples separately.

Model 5 comprised (for the 16-item version) the best fitting item loading that represented all five languages, i.e., items were constrained to have the same loading across the different languages.

Model 6 constrained the item loading to be the same in groups of surgical ($N = 487$) and conservatively-treated ($N = 705$) patients.

Model 7 constrained the item loading to be the same in groups of degenerative deformity ($n = 190$) and idiopathic deformity ($n = 615$) patients (for non-English versions of the questionnaire only, as this information was not available for the USA sample).

Model 8 constrained the item loading to be the same in groups of younger (aged ≤50 years) and older (>50 years) patients.

The next two models (Models 9 and 10) were longitudinal models including the data of 727 patients who had completed the SRS questionnaire at both baseline and follow-up. These models evaluated the stability of the factorial structure over time. Model 9 estimated loading for each item freely at baseline and follow-up whereas model 10 constrained the loading of each item to be the same at both baseline and follow-up.

Table 1 Demographic and baseline clinical data of the patients

Model	Total	Turkish	Spanish	French	German	English
N	1192	229	428	95	195	245
Age (years)	42 ± 19	31 ± 13	38 ± 15	50 ± 13	61 ± 19	60 ± 10
Gender (F:M; % F)	963:229 (81%)	159:70 (69%)	360:68 (84%)	76:19 (80%)	144:51 (74%)	224:21 (91%)
Height (m)	1.65 ± 0.09	1.65 ± 0.11	1.64 ± 0.09	1.65 ± 0.10	1.66 ± 0.09	–
Weight (kg)	63.0 ± 13.2	63.1 ± 15.4	60.7 ± 10.9	66.1 ± 14.1	66.6 ± 13.9	–
SRS subtotal score	3.4 ± 0.7	3.6 ± 0.6	3.5 ± 0.7	2.8 ± 0.6	3.2 ± 0.6	3.2 ± 0.5
Etiology						–
Idiopathic	623 (66%)	158 (69%)	345 (81%)	56 (59%)	64 (33%)	
Degenerative	190 (20%)	20 (9%)	29 (7%)	16 (17%)	125 (64%)	
Other	134 (14%)	51 (22%)	54 (12%)	23 (24%)	6 (3%)	
Treatment						
Surgical	487 (41%)	87 (38%)	109 (25%)	82 (86%)	82 (42%)	127 (52%)
Conservative	705 (59%)	142 (62%)	319 (75%)	13 (14%)	113 (58%)	118 (48%)

– Information not available in English language dataset

Cross-validation

A second large US sample of English-speaking patients who all received surgery was used for cross-validation of the models. Models 1, 2, 4, 7, 8, 9, and 10 were replicated with a sample of 772 patients who had completed the SRS-22 at baseline and 604 who had completed it at both baseline and follow-up.

Results

Table 2 shows the fit indicators and results of the nine CFA models that were tested. The one-factor model (Model 1) did not yield a good fit to the empirical data (RMSEA 0.14, CFI 0.65). The proposed four-factor structure with dimensions of function, pain, self-image, and mental health with five items each (Model 2) had reasonable fit parameters (RMSEA 0.09, CFI 0.87). Compared with model 1, model 2 had a significantly better fit as shown by the Chi squared difference test ($\Delta\chi^2(7) = 2544.61$, $p < 0.001$). In model 2a, item loading was estimated freely for the Spanish, Turkish, French, German, and English subsamples separately and the fit indices of that model approached acceptable levels (RMSEA 0.04). However, in model 3 with the restriction that item loading of each individual question should be the same across the five languages there was a significant decrease in the model fit (Model 3 compared with Model 2a: $\Delta\chi^2(64) = 261.80$, $p < 0.001$). Thus, this model comparison indicated non-equivalence of language versions.

The standardized item loading from the CFA of model 2a (unconstrained item loading estimated separately for the different language versions) is shown in Table 3. Model 2a revealed items that should be considered for omission (due to poor fit) in order to build a trimmed, optimized model of the SRS instrument; these items are marked bold in Table 2. The three items Q14 (self-image), Q15 (function), and Q17 (pain) consistently showed weak item loading within their factors across language versions and also in baseline and follow-up datasets (Table 2). All five "mental health" items showed reasonably high loading within their factor, but Q3 showed the lowest contribution and, in the interests of maintaining the same number of items for each factor, it was therefore chosen for removal. Other items such as Q8, Q10, and Q11 showed weak item loading in some language versions, but loading was not consistently low. Figure 1 shows the 16 questions that were retained (four per domain) in the trimmed 16-item version.

In model 4 the trimmed 16-item version of SRS was tested and achieved significantly better model fit than the 20-item version [Model 4 compared with Model 2: $\Delta\chi^2(64) = 746.51$, $p < 0.001$]. In model 4a, loading was estimated freely for the Spanish, Turkish, French, German, and English subsamples separately and the results showed significantly better fit in this trimmed 16-item version of SRS compared with the respective 20-item SRS [Model 4a compared with Model 2a: $\Delta\chi^2(320) = 897.33$, $p < 0.001$]. However, equivalence of language versions was still not reached in the trimmed 16-item SRS: comparison of model 5 (where the item loading was constrained to be the same across the five language versions) and model 4a showed a significant decrease of model fit [Model 5 compared with Model 4a: $\Delta\chi^2(48) = 202.88$, $p < 0.001$].

Equivalence was, however, good with respect to the type of treatment being received [surgical or conservative)

Table 2 Fit indicators and results of the confirmatory factor analysis models

Model	Model details	*N*	χ^2	*df*	χ^2/df	RMSEA	CFI	$\Delta\chi^2(df)$	*p*
Baseline									
Model 1	One-factor model (all 20 items measure the same construct), total sample	1192 Baseline	4183.91	169	24.76	0.14	0.65		
Model 2	Four factors (20 items with 5 items each on function, pain, self-image, and mental health), total sample	1192 Baseline	1639.30	162	10.12	0.09	0.87	Model 2 better than Model 1? Yes: 2544.61 (7)	<0.001
Model 2a	Same as Model 2 but with items estimated freely for Spanish (428), Turkish (229), French (95), German (195), and English (245) samples	1192 Baseline	2177.06	810	2.69	0.04	0.88		
Model 3	Same as Model 2 but with items constrained to have the same loading in Spanish (428), Turkish (229), French (95), German (195), and English (245) samples	1192 Baseline	2438.96	874	2.79	0.04	0.86	Model 3 as good as Model 2a? No, worse: 261.80 (64)	<0.001
Model 4	Trimmed model (16 items with 4 items each for function, pain, self-image, mental health; items 3, 14, 15, 17 excluded), total sample	1192 Baseline	892.79	98	9.11	0.08	0.92	Model 4 better than Model 2? Yes: 746.51 (64)	<0.001
Model 4a	Same as Model 4 but with items estimated freely for Spanish (428), Turkish (229), French (95), German (195), and English (245) samples	1192 Baseline	1279.73	490	2.61	0.04	0.91	Model 4a better than Model 2a? Yes: 897.33 (320)	<0.001
Model 5	Same as Model 4 but with items constrained to have the same loading in Spanish (428), Turkish (229), French (95), German (195), and English (245) samples	1192 Baseline	1482.61	538	2.68	0.04	0.90	Model 5 as good as Model 4a? No: 202.88 (48)	<0.001
Model 6	Same as Model 4 but with items constrained to have the same loading in surgical (487) and nonsurgical (705) patients	1192 Baseline	1000.42	208	4.81	0.06	0.91	Model 6 as good as Model 4? Yes: 107.63 (110)	0.546
Model 7	Same as Model 4 but with items constrained to have the same loading in degenerative (190) and idiopathic (622) patients	812 (no diagnosis information in English data)	755.19	208	3.63	0.06	0.91	Model 7 as good as Model 4 with 812 patients? Yes: 99.89 (110)	0.745
Model 8	Same as Model 4 but with items constrained to have the same loading in younger [≤50 years (634)] and older [>50 years (556)] patients	1192 Baseline	1045.98	208	5.03	0.06	0.90	Model 8 as good as Model 4? No: 153.19 (110)	0.004
Longitudinal									
Model 9	Same as Model 4, where loading of the same item can be different at baseline and follow-up (free estimation, trimmed model with 16 items in patients with baseline and follow-up data)	727	1393.04	432	3.22	0.06	0.93		
Model 10	Same as Model 4, but where loading of the same item has to be the same at baseline and follow-up (constrained, trimmed model with 16 items)	727	1410.99	444	3.18	0.05	0.93	Model 10 as good as Model 19? Yes: 17.95 (12)	0.117

Table 3 Standardized item loading of the confirmatory factor analysis of model 2a

Subscale	Item[a]	Response format	Cross-sectional sample (baseline)								Cross-validation sample	
			Turkish $N = 229$	Spanish $N = 428$	French $N = 95$	German $N = 195$	English $N = 245$	Total $N = 1192$	Baseline $N = 727$	FU $N = 727$	Baseline $N = 604$	FU $N = 604$
Function	Q5. What is your current level of activity?	Bedridden; primarily no activity; light labor and light sports; moderate labor and moderate sports; full activities without restriction[e]	0.73	0.57	0.80	0.70	0.68	0.65	0.68	0.64	0.80	0.68
	Q9. What is your current level of work/school activity?	100% normal, 75, 50, 25, 0% normal	0.71	0.65	0.68	0.70	0.74	0.68	0.73	0.73	0.73	0.67
	Q12. Does your back limit your ability to do things around the house?	Never, rarely, sometimes, often, very often	0.78	0.81	0.71	0.81	0.72	0.83	0.82	0.83	0.85	0.77
	Q15. Are you and/or your family experiencing financial difficulties because of your back?	**Severely, moderately, mildly, slightly, none**	**0.31**	**0.42**	**0.41**	**0.33**	**0.48**	**0.32**	**0.34**	**0.47**	**0.31**	**0.33**
	Q18. Does your back condition limit your going out with friends/family?[b]	Never, rarely, sometimes, often, very often	0.63	0.80	0.71	0.65	*0.47*	0.71	0.70	0.62	0.62	0.55
Pain	Q1. Which one of the following best describes the amount of pain you have experienced during the past 6 months?	None; mild; moderate; moderate to severe; severe	0.85	0.91	0.89	0.86	0.86	0.89	0.90	0.90	0.92	0.88
	Q2. Which one of the following best describes the amount of pain you have experienced over the last month?	None; mild; moderate; moderate to severe; severe	0.94	0.92	0.89	0.87	0.89	0.92	0.92	0.93	0.93	0.91
	Q8. Do you experience back pain when at rest?	Very often, often, sometimes, rarely, never?	0.61	0.65	0.63	*0.44*	*0.47*	0.60	0.60	0.66	0.62	0.72
	Q11. Which one of the following best describes your pain medication use for back pain?[c]	(None; non-narcotics weekly or less (e.g., aspirin, Tylenol, Ibuprofen); non-narcotics daily; narcotics weekly or less (e.g., Tylenol III, Percocet); narcotics daily	*0.34*	0.61	0.52	*0.47*	*0.42*	0.55	*0.46*	*0.51*	*0.53*	0.55

Table 3 continued

Subscale	Item[a]	Response format	Cross-sectional sample (baseline)								Cross-validation sample	
			Turkish $N = 229$	Spanish $N = 428$	French $N = 95$	German $N = 195$	English $N = 245$	Total $N = 1192$	Baseline $N = 727$	FU $N = 727$	Baseline $N = 604$	FU $N = 604$
	Q17. In the last 3 months, have you taken any days off of work, including household work, or school because of back pain?[d]	**0; 1; 2; 3; 4 or more days**	**0.42**	**0.50**	**0.51**	**0.32**	**0.32**	**0.41**	**0.37**	**0.35**	**0.38**	**0.34**
Self-image	Q4. If you had to spend the rest of your life with your back shape as it is right now, how would you feel about it?	Very happy; somewhat happy; neither happy nor unhappy; somewhat unhappy; very unhappy	0.51	0.65	0.57	0.67	*0.35*	0.63	0.62	0.70	0.42	0.68
	Q6. How do you look in clothes?	Very good; good; fair; bad; very bad	0.79	0.78	0.77	0.59	0.79	0.70	0.71	0.80	0.74	0.74
	Q10. Which of the following best describes the appearance of your trunk, defined as the human body except for the head and extremities?	Very good; good; fair; poor; very poor	0.82	0.79	*0.51*	*0.39*	0.69	0.70	0.69	0.84	0.66	0.78
	Q14. Do you feel that your back condition affects your personal relationships?	**None; slightly; mildly; moderately; severely**	**0.55**	**0.66**	**0.52**	**0.72**	**0.40**	**0.64**	**0.57**	**0.58**	**0.45**	**0.59**
	Q19. Do you feel attractive with your current back condition?	Yes, very; yes, somewhat; neither attractive nor unattractive; no, not very much; no, not at all	0.76	0.75	0.79	0.59	0.75	0.70	0.691	0.80	0.76	0.80
Mental health	**Q3. During the past 6 months have you been a very nervous person?**	**None of the time; a little of the time; some of the time; most of the time; all of the time**	**0.52**	**0.60**	**0.57**	**0.71**	**0.71**	**0.59**	**0.60**	**0.61**	**0.62**	**0.67**
	Q7. In the past 6 months have you felt so down in the dumps that nothing could cheer you up?	Very often; often; sometimes; rarely; never	0.78	0.79	0.90	0.88	0.77	0.81	0.79	0.83	0.77	0.78
	Q13. Have you felt calm and peaceful during the last 6 months?	All of the time; most of the time; some of the time; a little of the time; none of the time	0.73	0.66	0.45	0.72	0.68	0.66	0.66	0.73	0.79	0.82
	Q16. In the past 6 months, have you felt downhearted and blue?	Never; rarely; sometimes; often; very often	0.90	0.89	0.93	0.93	0.84	0.89	0.88	0.87	0.87	0.86

Table 3 continued

Subscale	Item[a]	Response format	Cross-sectional sample (baseline)								Cross-validation sample	
			Turkish $N = 229$	Spanish $N = 428$	French $N = 95$	German $N = 195$	English $N = 245$	Total $N = 1192$	Baseline $N = 727$	FU $N = 727$	Baseline $N = 604$	FU $N = 604$
	Q20. Have you been a happy person during the past 6 months?	None of the time; a little of the time; some of the time; most of the time; all of the time	0.68	0.70	0.54	0.67	0.83	0.69	0.71	0.75	0.76	0.76

Bold indicates items that were skipped from the 20-item SRS questionnaire to form the trimmed 16-item questionnaire. Italicized items indicate comparatively low loading when compared with loading of other items of this factor. The last four columns show the loading of a longitudinal model with baseline and follow-up loading for the 20 non-administrative items of the SRS-22 ($N = 720$ participants) and free estimation of loading (no constraint of item loading to be the same at baseline and follow-up)

[a] Wording is as per the official English SRS-22R version. http://www.srs.org/UserFiles/file/outcomes/srs-22_sample.pdf

[b] Previous formulation for Q18 in SRS-22: "Do you go out more or less than your friends?"; with responses being "much more, more, same, less, much less"

[c] Previous formulation for Q11 in SRS-22: "Which one of the following best describes your medication usage for your back?" (also gives the response option "Other (please specify)" with space to add the name of the medication and its usage (weekly or less, or daily)

[d] Previous formulation for Q17 in SRS-22: "In the past 3 months, have you taken any sick days from work/school due to back pain and, if so, how many?"

[e] Previous response option for Q5 in SRS-22: Bedridden/wheelchair; primarily no activity; light labor, such as household chores; moderate manual labor and moderate sports, such as walking and biking; full activities without restriction

[Model 6 compared with Model 4: $\Delta\chi^2(110) = 107.63$, $p = 0.546$]. In addition, equivalence was good with respect to the etiology of the deformity (idiopathic or degenerative) [Model 7 compared with Model 4: $\Delta\chi^2(110) = 99.89$, $p = 0.745$]. Equivalence was less good with respect to patient age. Comparison of model 8 with model 4 showed a significant difference in χ^2 [$\Delta\chi^2(110) = 153.19$ (110), $p = 0.004$].

The trimmed 16-item model using the data of the 727 patients available at both baseline and the one-year follow-up yielded an acceptable fit to the empirical data (model 9; RMSEA 0.06, CFI 0.93). Moreover, the constraint that individual items should have the same loading at baseline and follow-up did not decrease model fit [Model 10 compared with Model 9: $\Delta\chi^2(12) = 17.95$, $p = 0.117$]. Thus, equivalence of item loading and factor structure over time was confirmed, yielding a model with acceptable fit (RMSEA 0.05, CFI 0.93; see Fig. 2 for item loading).

Cross-validation

The data from the patients in the US sample used for cross-validation showed item loading that was very similar to those of the other samples. Moreover, the decisions regarding item selection were confirmed in the cross-validation sample: consistent with the other data, Q15, Q17, Q14, and Q3 showed the weakest loading within the four five-item subscales (cf. Table 2 last two columns). The fit indicators of Models 1, 2, 4, 7, 8, 9, and 10 are shown in Table 4. Again, omission of Q15, Q17, Q14, and Q3 significantly increased the model fit. In model 4, the trimmed 16-item version of SRS was tested and achieved significantly better model fit than the 20-item version [Model 4 compared with Model 2: $\Delta\chi^2(64) = 583.25$, $p < 0.001$]. Similar to the previous analyses, equivalence was good with respect to the etiology of deformity (idiopathic or degenerative) [Model 7 compared with Model 4: $\Delta\chi^2(110) = 114.72$, $p = 0.360$] but less good with respect to the age of the patients; comparison of model 8 with model 4 showed a significant difference in $\chi^2(\Delta\chi^2(110) = 143.29$ (110), $p = 0.018$). The trimmed 16-item model using the data available at both baseline and one-year follow-up for 604 patients yielded an acceptable fit to the empirical data (model 9). However, the constraint that individual items should have the same loading at baseline and follow-up did decrease model fit [Model 10 compared with Model 9: $\Delta\chi^2(12) = 87.26$, $p < 0.001$]. Thus, equivalence of item loading and factor structure across time was not confirmed in the US cross-validation sample. The considerably lower loading of Q4 at baseline than at follow-up (Table 2) likely contributed to this result. The fit of model 10 was, however, nonetheless good (RMSEA 0.05, CFI 0.93).

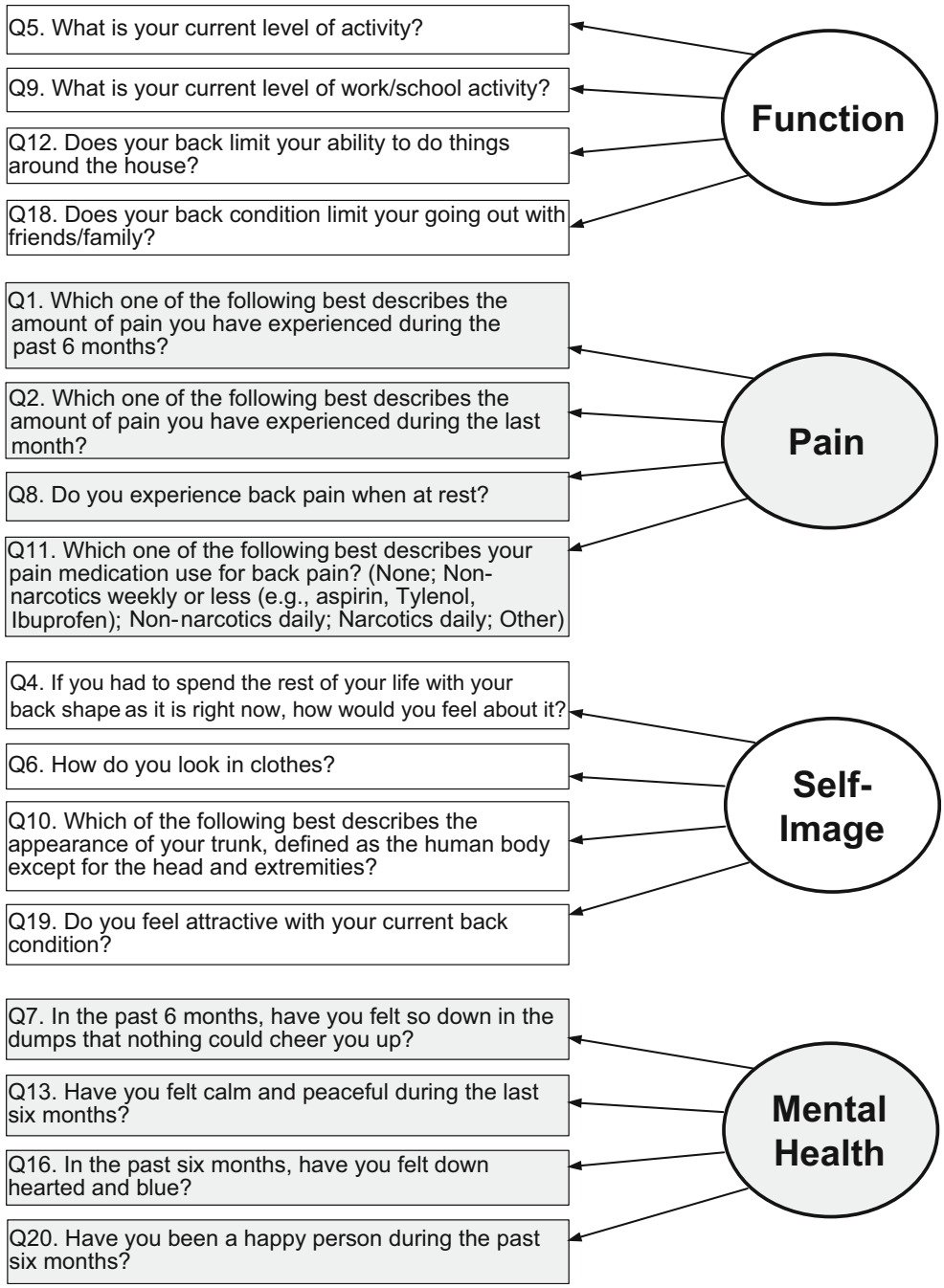

Fig. 1 Trimmed model including the 16 remaining questions and constraining the structure to be the same in all four languages (good fit to the baseline data: Model 5 in Table 2, RMSEA = 0.04, CFI = 0.90)

Discussion

In a recent systematic review, Monticone et al. [10] described the psychometric properties of 17 published translations of the SRS-22 instrument. The absence of any description of the cross-cultural validation procedures, for any of the languages investigated, prevented the authors from understanding whether the constructs underlying the original instrument were adequately reflected in the translations [10]. Further, the 4-factor structure of the SRS-22, originally designed for patients with AIS, has not been evaluated in any language in patients with ASD. The present study aimed to rectify this situation, by performing confirmatory factor analysis of the 4-factor structure on English, Spanish, German, French, and Turkish versions of the SRS-22 in patients with ASD. We showed that, although the 4-factor structure was upheld, the fit of the items was not perfect.

Over the last two decades, a lot of work has gone into developing the SRS-outcome instrument and cross-culturally adapting it in many different languages [10]. Of all the instruments used in the assessment of patients with ASD, the SRS-22 is the one that has been subjected to the most thorough and extensive evaluation of its clinimetric qualities [18]. Numerous clinical studies have used it as their primary outcome measure [19]. For these reasons, we chose to view the problems identified with the instrument as the impetus for improving it (see later) rather than an indication of the need for an entirely new instrument. This was expected to promote continuity in the interpretation of data collected in past and future studies. Few other outcome instruments have been subjected to scrutiny of their measurement properties across their different language versions [20], and even fewer have been shown to satisfy the quality criteria [21] in relation to their factor structure; the relatively minor and easily-rectifiable deficiencies in the SRS-22 must hence be viewed from this perspective.

The 20 non-management items of the SRS-22 have different response formats in relation to what they provide information about: frequency, intensity, days, percentages, types of medication, etc. This can be a problem in factor analysis, because some error relates specifically to the format itself. With differing response formats the comparability between items is often lower, making it less likely that items which are expected to load together on a common factor will actually do so. Given that the response formats differ so much across the 20 items of the SRS-22, the four-factor SRS had a surprisingly good fit.

When estimates were compared across the five different languages, there was no language version that obviously "did not work", in the sense that the relative item loading was very different from the item loading for the other languages. Nonetheless, some items showed slightly weaker loading in some language versions than in others. The three items that consistently showed weak item loading within their factors, across all language versions of the instrument and in both baseline and follow-up datasets, were Q14, Q15, and Q17 (Fig. 1). Question 14 (Do you feel that your condition affects your personal relationships?) in the "self-image" domain, is somewhat ambiguous and it is perhaps not clear as to what is intended by the term "personal relationships". Previous exploratory factor analyses of the data of patients with AIS have observed that this item often loads (more) strongly on the function domain [10, 13, 22, 23], perhaps because the typical adolescent patient interprets it as referring to the ability to meet friends and carry out normal activities with friends and family [13]. It was suggested that in an adult population the item may instead target the interference of the back condition with intimate relations, which is closer to the Self-image concept [13]. The results of the present study in adult patients do not, however, support this argument.

Question 15 (Are you and/or your family experiencing financial difficulties because of your back?) in the "function" domain, has repeatedly been identified as problematic in other studies of the SRS [5, 14] and has previously been recommended for removal [14]. It is somewhat unclear why it should be included in a domain assessing "function", and it would also appear to have little relevance to the present-day patient, especially in countries with the necessary social security support to acquire orthoses, obtain treatment, medical care, disability allowance, etc.

With respect to item 17 (In the past 3 months, have you taken any sick days from work/school due to back pain and, if so, how many?) it might be questioned why this was chosen for inclusion in the "pain" domain. Being unable to perform normally and hence taking time off work tends to be more of a "behavioral" or "functional" issue and might be expected to better represent the "function" domain. A previous exploratory factor analysis of the Japanese version of the SRS-22 found that this item loaded more heavily on "function" than "pain" [24]. Another flaw of this item lies in its response scale, in that the distance between responses is not proportional (i.e., the difference between "1" and "2" days, and "2" and "3" days, is not the same as between "3" and "4 or more" days, where the "more" could be up to 3 months). A response scale similar to that used for other disability items (e.g., as recommended by Deyo et al. [25] and subsequently used as a single item in the Core Outcome Measures Index [26, 27]) might have been more appropriate here. And finally, in the official English version of the SRS-22R (http://www.srs.org/professionals/online-education-and-resources/patient-outcome-questionnaires), item 17 states "work, including housework", but the word "housework" was not included in any of the language versions investigated here, presumably because its introduction into the SRS-22R was not explicitly mentioned in the paper describing the revision [5], and hence the older formulation without "housework" was retained.

Based on the aforementioned shortcomings of questions 14, 15 and 17, we opted to remove them from the original set of 20 non-management items. In order to maintain the structure of four domains, with an equal number of items representing each, we also elected to discard the lowest-loading item from the "mental health" domain (item 3). The resulting trimmed, 16-item model performed significantly better than the 20-item model. For the 16-item model, there was equivalence of item loading within the four-factor structure across treatment type (surgical or conservative), aetiology (degenerative or idiopathic ASD) and time of assessment (baseline and follow-up). However, even the 16-item instrument—although performing

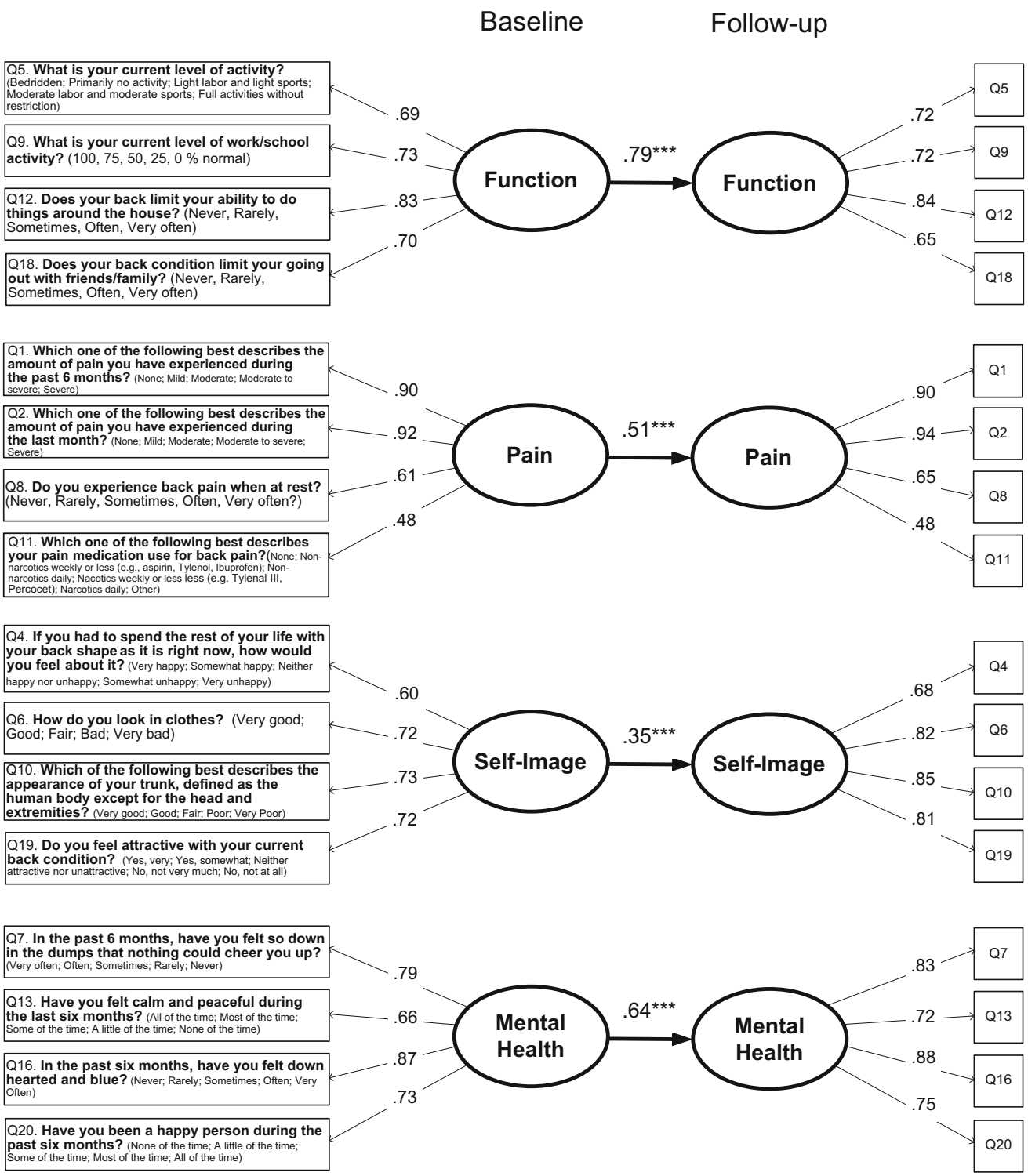

Fig. 2 Trimmed model including the 16 remaining questions and constraining the structure to be the same at baseline and follow-up (good fit to the data: Model 10 in Table 2, RMSEA = 0.05, CFI = 0.93)

significantly better than the 20-item version—still did not show equivalence in item loading across languages or age groups (although the age comparison was also confounded by language, because age-levels differed across the language samples; Table 1). This suggests that further work is still required to improve the comparability of the wording of these 16 items, and to ensure their content is relevant to patients of all ages. Some items were

Table 4 Fit indicators for Models 1, 2, 4, 7, 8, 9, and 10 in the cross-validation sample

Model		*N*	χ^2	*df*	χ^2/df	RMSEA	CFI	$\Delta\chi^2$ (*df*)	*p*
Baseline									
Model 1	One-factor model (all 20 items measure the same construct), total sample	772 Baseline	2886.17	169	17.08	0.14	0.64		
Model 2	Four factors (20 items with 5 items each on function, pain, self-image, and mental health), total sample	772 Baseline	1056.41	162	6.52	0.08	0.88	Model 2 better than Model 1? Yes: 1829.76 (7)	<.001
Model 4	Trimmed model (16 items with 4 items each for function, pain, self-image, mental health; items 3, 14, 15, 17 excluded), total sample	772 Baseline	473.16	98	4.83	0.07	0.94	Model 4 better than Model 2? Yes: 583.25 (64)	<.001
Model 7	Same as Model 4 but with items constrained to have the same loading in degenerative (129) and idiopathic (478) patients	607 Baseline	523.01	208	2.51	0.05	0.93	Model 7 as good as Model 4 with 607 patients? Yes: 114.72 (110)	0.360
Model 8	Same as Model 4 but with items constrained to have the same loading in younger [≤50 years (248)] and older [>50 years (524)] patients	772 Baseline	616.45	208	2.96	0.05	0.93	Model 8 as good as Model 4? No: 143.29 (110)	0.018
Longitudinal									
Model 9	Same as model 4, where loading of the same item can be different at baseline and follow-up (free estimation, trimmed model with 16 items)	604	1110.58	432	2.57	0.05	0.94		
Model 10	Same as Model 4, but where loading of the same item has to be the same at baseline and follow-up (constrained, trimmed model with 16 items)	604	1197.84	444	2.70	0.05	0.93	Model 10 as good as Model l 9? No: 87.26 (12)	<0.001

χ^2 Chi-square value indicates the minimum discrepancy between empirical covariance structures and those implied by the model, *df* degrees of freedom, χ^2/df minimum discrepancy divided by its degrees of freedom, as an indicator of fit, *p p* value of minimum discrepancy divided by its degrees of freedom, which should be smaller than 2 [17], *CFI* comparative fit index

CFI higher than 0.90 in the mediation model reflects acceptable fit between the model and the data [17], RMSEA value below 0.05 reflects a good fit of the model [17]. To test the fit between two nested models the difference in χ^2 and *df* $\Delta\chi^2(df)$ was calculated (χ^2 difference test). *p* indicates a significantly better fit of the model with lower χ^2 value

particularly weak in some of the languages (Q4 (rest of life with current back shape) in English; Q8 (back pain at rest) in German and English; Q10 (appearance of the trunk) in German and French; and Q11 (medication usage) in German, English and Turkish; see Table 3), and the wording of these items should be double-checked and where necessary improved or made more consistent for those language versions. Q11 in particular seems to differ in relation to the complexity of the language used in the response options ("strong/weak painkillers" versus "narcotics/non-narcotics"), and whether specific examples of types of medication are given (no examples in Spanish, different examples of non-narcotics and narcotics in the other languages). Other items that appeared to fit reasonably well, but nonetheless clearly enquired about different concepts in the different languages should also be addressed [e.g. item 18 (going out with friends/family), which used the SRS-22 "R" wording in Spanish and Turkish but not in English, German or French]. Part of these discrepancies may have arisen because revision of the English version of the SRS-22 to produce the SRS-22R was not accompanied by the simultaneous revision of other language versions. If no SRS-22R version for the given language was found in the literature at the start of our study (as was the case for French and German in the ESSG, and for the US English data too), then the wording of the items from the original SRS-22 (or its predecessor, the SRS-30) was used. Complicating matters further, it would seem that not all the amendments made in creating the current "SRS-22R" (http://www.srs.org/UserFiles/file/outcomes/srs-22_sample.pdf) have been specifically documented in the publication describing the revision [5]. For example, as mentioned above, the addition of the word "housework" to item 17 in the current SRS-22R does not appear to be documented anywhere, and the specification of "pain medication for back pain" (item 11, medication) is simply mentioned in passing in the discussion of Lai et al. [22]. Consequently, neither of these two modifications were included in any of

Table 5 Suggestions for improving the current quality of the different language versions of the SRS-22 evaluated in the present study

Item number	Domain	Item content (in brief)	Recommendation
1	Pain	Pain 6 months	No change
2	Pain	Pain 1 month	No change
3	Mental health	Nervous	Remove
4	Self-image	SSWB shape	Add "shape" to Spanish version
5	Function	Current activity	Remove "Wheelchair" from response 1 in German and Turkish; add "light sports" to response 2 and remove sporting examples from response 4 in German, French, and Turkish
6	Self-image	How you look?	No change
7	Mental health	Down in dumps	No change
8	Pain	LBP rest	No change
9	Function	Current activity work/school	Ensure school is included in the question (in some French versions, this might be missing); improve language in German version—"Auf welcher Ebene" is not native German sounding.
10	Self-image	Appearance trunk	No change
11	Pain	Medication use	Make sure question is "PAIN medication use for back PAIN" in all languages (add the words PAIN in each place). Give examples of the types of medication in each category (narcotic/non-narcotic) in Spanish version. Perhaps add the option for "other" and the frequency of usage, for patients who do not know what category their pain medication falls into. The examiner can then code it accordingly.
12	Function	Ability household	Improve the language in German version (to make it clear it is referring to household jobs not just things you do when you are at home; "doing things around the house" would imply housework, but not for example sitting watching TV)
13	Mental Health	Calm and peaceful	No change
14	Self-image	Personal relationships	Remove
15	Function	Financial problems	Remove
16	Mental health	Downhearted	No change
17	Pain	Sick days work/school	Remove
18	Function	Going out with friends/family	Develop 22-R version of this item (does your back condition limit your going out with friends/family) and its corresponding response options in German and French
19	Self-image	Feel attractive?	Improve the language in the German version ("Zustand Ihres Ruckens" does not sound native; use "Ruckenerkrankung"?)
20	Mental health	Happy?	No change
21[a]	Satisfaction	Satisfied with results of back management	Add "back" to Spanish version. Remove "keine Behandlung" from 3rd response option of German versions (a "not applicable" option is NOT the same as a mid-scale response to the question)
22[a]	Satisfaction	Have same management again?	Change "wahrscheinlich" to "ja, wahrscheinlich" for 2nd response option in German version; remove "keine Behandlung" from 3rd response option of German versions (a "not applicable" option is NOT the same as a mid-scale response to the question)

[a] Not formally evaluated in the present study, but apparent discrepancies indicated here

the language versions used in the present study (even the Spanish and Turkish, which otherwise used the official SRS-22R formulation). For item 11, all versions simply enquired about "medication usage for your back", and this may explain its generally low loading on the pain domain in all languages. Future iterations of the instrument, specifying pain medication for back pain, may improve this item's loading. Similarly, the official Spanish version of the SRS-22R (http://www.srs.org/UserFiles/file/outcomes/srs-22_spanish.pdf) does not include the word "shape" in item 4, in enquiring how one would feel about spending the rest of one's life with the back (shape) one has now. This

was because the version that was used to create the Spanish version (published in the Appendix of Asher et al. [2]), item 4 did not contain the word shape (despite the fact that "shape" was in the original longer SRS-30 versions). Only in the introduction to their article in 2003 [3] was the addition of the word "shape" to item 4 and its inclusion in the self-image (rather than pain) domain mentioned. Even though this item showed an adequate fit in the Spanish data, the discrepancy should be amended in the official Spanish version of the instrument to improve comparability with the other languages. A summary of all our recommended changes is shown in Table 5.

In their systematic review, Monticone et al. [28] advised that, following confirmatory factor analyses to verify the structure of the cross-cultural adaptations of the SRS-22, more extensive studies of their psychometric properties (e.g., responsiveness) should be carried out. The present study suggests that this should perhaps be done using the shorter, more structurally valid version of the instrument shown in Fig. 1. Studies of the test–retest reliability of this shorter version as a stand-alone instrument in each language should also be carried out, to ensure that the reduction in the number of items does not serve to threaten this psychometric property of the instrument.

The present study has a number of weaknesses that must be acknowledged. First, although we used the official versions of the SRS instrument in the different languages, in carefully comparing the wording it became apparent that there were discrepancies in the exact expressions used as well as inaccuracies in the translations. This was partly the result of "revisions" not being clearly detailed in the literature or introduced in all languages simultaneously (see earlier), but also partly due to deficiencies in the cross-cultural adaptations themselves. The same may well apply to other spine outcome instruments that have been translated into many languages; few have been subjected to the level of scrutiny applied here that would allow us to draw comparisons [20]. Second, the data were extracted from an existing database rather than collected prospectively for the given research question, and the characteristics of the patients differed somewhat across the languages, particularly regarding aetiology of the deformity and, hence, age. Although no restrictions have been published regarding the applicability of the instrument depending on age-group, aetiology, disease severity, etc., further studies should evaluate whether equivalence in item loading across languages is improved when these factors are held constant.

In conclusion, we recommend removal of the worst-fitting item in each of the four domains of the SRS-22 (items 3, 14, 15, 17), together with adaptation and standardization of the other items across language versions, to provide an improved version of the SRS instrument. This would include the 16 non-management items plus the two management items, forming a new SRS-18. We would hope that, in the fullness of time, and if shown to be reliable and responsive to the effects of treatment, the SRS-18 would become the commonplace instrument of choice in patients with ASD. Algorithms could be generated to convert the scores derived from one version to those for another version (as described by [29]), to allow comparisons across different versions for each of the languages.

Acknowledgements Funding was provide by Depuy Synthes Spine research Grant.

Compliance with ethical standards

Conflict of interest None of the authors has any potential conflict of interest.

References

1. Haher TR, Gorup JM, Shin TM, Homel P, Merola AA, Grogan DP, Pugh L, Lowe TG, Murray M (1999) 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) 24:1435–1440
2. Asher MA, Min Lai S, Burton DC (2000) Further development and validation of the Scoliosis Research Society (SRS) outcomes instrument. Spine (Phila Pa 1976) 25:2381–2386
3. Asher MA, Min Lai S, Burton D, Manna B (2003) Scoliosis Research Society-22 Patient questionnaire- responsiveness to change associated with surgical treatment. Spine 28:70–73
4. Asher M, Min Lai S, Burton D, Manna B (2003) The reliability and concurrent validity of the Scoliosis Research Society-22 Patient questionnaire for idiopathic scoliosis. Spine (Phila Pa 1976) 28:63–69. doi:10.1097/01.BRS.0000047634.95839.67
5. Asher MA, Lai SM, Glattes RC, Burton DC, Alanay A, Bago J (2006) Refinement of the SRS-22 health-related quality of life questionnaire function domain. Spine (Phila Pa 1976) 31:593–597. doi:10.1097/01.brs.0000201331.50597.ea00007632-200603010-00018
6. Jain A, Sponseller PD, Negrini S, Newton PO, Cahill PJ, Bastrom TP, Marks MC, Harms Study G (2015) SRS-7: a valid, responsive, linear, and unidimensional functional outcome measure for operatively treated patients with AIS. Spine (Phila Pa 1976) 40:650–655. doi:10.1097/BRS.0000000000000836
7. Caronni A, Zaina F, Negrini S (2014) 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 35:784–799
8. Rothenfluh DA, Neubauer G, Klasen J, Min K (2012) Analysis of internal construct validity of the SRS-24 questionnaire. Eur Spine J 21:1590–1595. doi:10.1007/s00586-012-2169-3
9. Jain A, Lafage V, Kelly MP, Hassanzadeh H, Neuman BJ, Sciubba DM, Bess S, Shaffrey CI, Ames CP, Scheer JK, Burton D, Gupta MC, Hart R, Hostin RA, Kebaish KM (2016) Validity, reliability, and responsiveness of SRS-7 as an outcomes assessment instrument for operatively treated patients with adult spinal deformity. Spine (Phila Pa 1976) 41:1463–1468. doi:10.1097/BRS.0000000000001540
10. Monticone M, Baiardi P, Calabro D, Calabro F, Foti C (2010) Development of the Italian version of the revised Scoliosis Research Society-22 Patient questionnaire, SRS-22r-I. Spine 35:1412–1417

11. Bago J, Climent JM, Ey A, Perez-Grueso FJ, Izquierdo E (2004) The Spanish version of the SRS-22 Patient questionnaire for idiopathic scoliosis: transcultural adaptation and reliability analysis. Spine (Phila Pa 1976) 29:1676–1680
12. Climent JM, Bago J, Ey A, Perez-Grueso FJ, Izquierdo E (2005) Validity of the Spanish version of the Scoliosis Research Society-22 (SRS-22) Patient questionnaire. Spine (Phila Pa 1976) 30:705–709
13. Beauséjour M, Joncas J, Goulet L, Roy-Beaudry M, Parent S, Grimard G, Forcier M, Lauriault S, Labelle H (2009) Reliability and validity of adapted French Canadian version of Scoliosis Research Society Outcomes questionnaire (SRS-22) in Quebec. Spine 34:623–628
14. Alanay A, Cil A, Berk H, Acaroglu E, Yazici M, Akcali O, Kosay C, Genc Y, Surat A (2005) Reliability and validity of adapted Turkish version of Scoliosis Research Society-22 (SRS-22) questionnaire. Spine 30:2464–2468
15. Niemeyer T, Schubert C, Halm HF, Herberts T, Leichtle C, Gesicki M (2009) Validity and reliability of an adapted german version of Scoliosis Research Society-22 questionnaire. Spine (Phila Pa 1976) 34:818–821. doi:10.1097/BRS.0b013e31819b33be
16. Pellise F, Vila-Casademunt A, Ferrer M, Domingo-Sabat M, Bago J, Perez-Grueso FJ, Alanay A, Mannion AF, Acaroglu E (2015) Impact on health related quality of life of adult spinal deformity (ASD) compared with other chronic conditions. Eur Spine J 24:3–11. doi:10.1007/s00586-014-3542-1
17. Schermelleh-Engel K, Moosbrugger H, Müller H (2003) Evaluating the fit of structural equation models: test of significance and descriptive goodness-of-fit measures. Methods Psychol Res Online 8:23–74
18. Faraj SSA, van Hooff ML, Holewijn RM, Polly DW Jr, Haanstra TM, de Kleuver M (2017) Measuring outcomes in adult spinal deformity surgery: a systematic review to identify current strengths, weaknesses and gaps in patient-reported outcome measures. Eur Spine J. doi:10.1007/s00586-017-5125-4
19. Guzman JZ, Cutler HS, Connolly J, Skovrlj B, Mroz TE, Riew KD, Cho SK (2016) Patient-reported outcome instruments in spine surgery. Spine (Phila Pa 1976) 41:429–437. doi:10.1097/BRS.0000000000001211
20. Costa LO, Maher CG, Latimer J (2007) Self-report outcome measures for low back pain: searching for international cross-cultural adaptations. Spine 32:1028–1037
21. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60:34–42
22. Lai SM, Asher MA, Burton DC, Carlson BB (2010) Identification of Scoliosis Research Society-22r Health-Related Quality of Life questionnaire domains using factor analysis methodology. Spine (Phila Pa 1976) 35:1236–1240. doi:10.1097/BRS.0b013e3181dbdb38
23. Potoupnis M, Papavasiliou K, Kenanidis E, Pellios S, Kapetanou A, Sayegh F, Kapetanos G (2012) Reliability and concurrent validity of the adapted Greek version of the Scoliosis Research Society-22r questionnaire. A cross-sectional study performed on conservatively treated patients. Hippokratia 16:225–229
24. Hashimoto H, Sase T, Arai Y, Maruyama T, Isobe K, Shouno Y (2007) Validation of a Japanese version of the Scoliosis Research Society-22 Patient questionnaire among idiopathic scoliosis patients in Japan. Spine (Phila Pa 1976) 32:E141–E146. doi:10.1097/01.brs.0000255220.47077.33
25. Deyo RA, Andersson G, Bombardier C, Cherkin DC, Keller RB, Lee CK, Liang MH, Lipscomb B, Shekelle P, Spratt KF, Weinstein JN (1994) Outcome measures for studying patients with low back pain. Spine 19:2032S–2036S
26. Mannion AF, Elfering A, Staerkle R, Junge A, Grob D, Semmer NK, Jacobshagen N, Dvorak J, Boos N (2005) Outcome assessment in low back pain: how low can you go? Eur Spine J 14:1014–1026
27. Mannion AF, Porchet F, Kleinstück F, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective: Part 1. The Core Outcome Measures Index (COMI) in clinical practice. Eur Spine J 18:367–373
28. Monticone M, Nava C, Leggero V, Rocca B, Salvaderi S, Ferrante S, Ambrosini E (2015) Measurement properties of translated versions of the Scoliosis Research Society-22 Patient questionnaire, SRS-22: a systematic review. Qual Life Res 24:1981–1998. doi:10.1007/s11136-015-0935-5
29. Lai SM, Burton DC, Asher MA, Carlson BB (2011) Converting SRS-24, SRS-23, and SRS-22 to SRS-22r: establishing conversion equations using regression modeling. Spine (Phila Pa 1976) 36:E1525–E1533. doi:10.1097/BRS.0b013e3182118adf