**REVIEW ARTICLE**

# Evaluating measures of quality of life in adult scoliosis: a systematic review and narrative synthesis

James E. Archer[1] · Charles Baird[1] · Adrian Gardner[1,3] · Alison B. Rushton[2] · Nicola R. Heneghan[3]

## Abstract

**Purpose** To systematically review and synthesise the evidence on the measurement properties of patient-reported outcome measure (PROMs) used to assess the quality of life in patients with adult scoliosis.

**Method** Based on the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines and a published protocol, a two-stage search was conducted and reported in line with the Preferred Reporting Items for Systematic review and Meta-analysis (PRISMA). Stage one identified all studies of patients with adult scoliosis which included PROMs of health-related quality of life (HR-QOL). Databases including AMED, CINAHL, EMBASE, Medline, PsychINFO and Pubmed were searched from inception until 31st December 2020. This derived list of PROMs, was then utilised for a stage 2 search to identify studies which evaluated the measurement properties of the PROMs. Two reviewers independently performed the searches, study screening, selection and risk of bias assessment using the COSMIN tool. The overall quality of the evidence was assessed using a modified Grading of Recommendations Assessment, Development and Evaluation approach.

**Results** Stage one yielded 16 PROMs of HR-QOL with the Oswestry Disability Index (ODI) and Scoliosis Research Society-22 (SRS-22) the most used. Stage two identified three stage one PROMs that fulfilled eligibility criteria: SRS-22, ODI and SRS-22r; with five studies investigating measurement properties in an adult scoliosis population. The SRS-22 was the most comprehensively evaluated PROM in this cohort with very low-quality evidence indicating indeterminate reliability, sufficient construct validity and sufficient responsiveness of the SRS-22. There is very low-quality evidence indicating sufficient responsiveness of the ODI. There is very low-quality evidence indicating indeterminate cross-cultural validity for the SRS-22r. All other measurement properties in the SRS-22, ODI and SRS-22r have not been evaluated.

**Conclusion** A large number of PROMs are being utilised in the adult scoliosis population and of these, the most commonly utilised are the ODI and SRS-22. The SRS-22, ODI and SRS-22r are the only PROMs to have had their measurement properties evaluated in the adult scoliosis population. The findings of this systematic review are that there currently is not sufficient evidence on the measurement properties of any PROMs in adult scoliosis. Further research is now urgently required to assess the measurement properties of these PROMs.

**Keywords** Adult scoliosis · Health-related quality of life · Patient-reported outcome measures · Systematic review · Measurement properties

✉ Adrian Gardner
adrian.gardner@nhs.net

Extended author information available on the last page of the article

## Introduction

Health-related quality of life (HR-QOL) is defined by the World Health Organisation as 'an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns' [1]. This is a broad definition, which takes into account an array of important elements of which health and disability form an important part. However, even defining disability can be challenging.

The International Classification of Functioning, Disability and Health [2] aims to formulate a structure for both the medical and social models of disability, defining how these, along with an individual's perceptions, form disability. Scoliosis impacts heavily on these models as it changes an individual's health perceptions, potentially leading to disability.

Adult scoliosis (AS) is a condition requiring different considerations when comparing to other forms of scoliosis. Individuals with AS experience an array of symptoms including back pain and focal neurology [3]. Furthermore, the AS population represents a wide age-range and spectrum of co-morbidities when compared to adolescent idiopathic scoliosis (AIS). The aetiology of AS is also different to AIS, being either degenerative or idiopathic in nature. All these factors mean that individuals often report a significant impact on their HR-QOL which has been classed as more significant than other health conditions such as heart failure, chronic lung disease or diabetes [4].

The reported incidence of AS varies in the literature from 2.9 to 32% [5–8]. The incidence of AS, and specifically adult degenerative scoliosis, is likely to increase with an ageing population [9]. Management of AS is aimed at improving HR-QOL through improving pain and function [10–12], with a lesser focus on the cosmetic appearance of the spine and torso, which is more of an issue in those with AIS [13].

Conservative measures for managing AS can be ineffective due to the degenerative nature of the condition [14]. Furthermore, the understanding and management of AS has undergone a paradigm shift over the last 10 years, with an increasing awareness and subsequent increased demand for active management [15, 16] leading to an increase in surgical management. Health economic evaluation of surgical interventions have been shown to be cost-effective [17]. However, to be able to accurately assess the impact of any intervention in this specific cohort of individuals, there is a requirement for appropriate patient-reported outcome measures (PROMs). PROMs have a well-established role within the scoliosis population as a whole [18]. The most widely used PROM in individuals with scoliosis is the Scoliosis Research Society questionnaire (SRS-22) which has been adapted and modified over time [19–22]. However, in the AS literature, a wide array of PROMs are also used, including the Oswestry Disability Index (ODI), the SRS-22 and Short Form-36 (SF-36).

It is key, however, that the measurement properties of any PROM that is used in any group of individuals is accurate and appropriate [23]. Performing research with PROMs without establishing their measurement properties is a waste of research resources [24] and provides little value. The measurement properties for PROMs used in AS have not been reported in the literature. It is, therefore, essential that this assessment is performed to identify and reduce the risk of bias and inaccuracy in any results reported using PROMs

[25] in this group of individuals. This will allow a full understanding of how best to assess HR-QOL and subsequent management in this discrete patient population.

## Objective

To systematically review and synthesise the evidence on the measurement properties of patient-reported outcome measure (PROMs) used to assess the quality of life in patients with adult scoliosis.

## Methods

### Design

This review and synthesis of evidence followed a previously published protocol [26] and is registered with PROSPERO (CRD42020219437). The review was designed using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology [25], which aims to improve the selection of the best outcome measures in both clinical practice and research. The results of the review are reported using the preferred reporting items for systematic review and meta-analysis (PRISMA) guidance [27]. This guidance aims to ensure all systematic reviews and meta-analyses report a minimum set of items with the overall aim of improving the quality and the usefulness of systematic reviews. The PRISMA checklist is included as supplementary file 1.

### Search strategy

The searches were performed in two stages. In stage one, any PROMs used to report on HR-QOL in individuals with AS were identified. This was completed to identify the frequency of use of the various PROMs and to help inform the search strategy for stage two. In stage two, studies assessing the measurement properties of the PROMs identified in stage one were identified. Stage two searches were grouped according to whether they were generic or disease-specific PROMs.

### Eligibility criteria: stage one

#### Inclusion criteria

1. Individuals diagnosed with AS (> 10° Cobb angle [28] with an age of 18 years or more)
2. Any study which included a PROM of HR-QOL, as defined by the World Health Organisation: "Individuals' perceptions of their position in life in the context

of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns" [1]. PROMs were defined using the Core Outcome Measures in Effectiveness Trials (COMET) taxonomy [29]

## Eligibility criteria: stage two

### Inclusion criteria

1. Individuals diagnosed with AS (> 10° Cobb angle [28] with an age of 18 years or more).
2. Any study which evaluated the measurement properties (reliability, internal consistency, measurement error, validity, content validity, structural validity, interpretability, responsiveness) of the PROMs identified in stage one.

The exclusion criteria for both stages of the search were studies not in the English language and previous systematic reviews.

## Data sources

A comprehensive search strategy was performed using the National Institute for Health and Care Excellence (NICE) Healthcare Databases Advanced Search (HDAS) tool. The following databases were searched: AMED, CINAHL, EMBASE, Medline, PsychINFO and Pubmed from inception to 31st December 2020.

The details of the searches are included as supplementary files 2 and 3.

## Study selection

A standardised blinded selection process was performed by two authors (JA and CB) independently. In stages one and two, the titles and abstracts were assessed against the predetermined eligibility criteria. For any study where the title and abstract was not clear with regards the inclusion and exclusion criteria, the full-text study was retrieved and reviewed. After unblinding, any disagreements were discussed and consensus agreed between the two authors. In the event of continued disagreement, the third reviewer (AG) acted as an arbitrator.

A PRISMA flow diagram was constructed to demonstrate this process [27, 30] for each stage of the search (Figs. 1, 2 and 3).

## Data extraction

Included studies had data extracted on the study characteristics, participant characteristics, PROM and measurement properties (Table 1).

## Risk of bias assessment

The COSMIN checklist was used to assess for the risk of bias (ROB) of each study as one of the assessments of the quality of the paper. Each item describing a measurement property was rated using the scale (very good, adequate, doubtful or inadequate). Two reviewers (JA and CB) assessed study risk of bias independently with disagreements resolved by discussion. In the event of continued disagreement, the third reviewer (AG) acted as an arbitrator. This step is important within the COSMIN process as it helps to dictate what overall rating a study receives and whether the rating needs to be 'downgraded due to risk of bias'.

## Synthesis of data

Data were examined to assess whether a meta-analysis was appropriate. In the absence of a large number of studies or homogeneous data precluding meta-analysis, a narrative synthesis was performed. The different measurement properties for each PROM were rated as 'sufficient', 'insufficient' or 'indeterminate' as directed by the COSMIN process [25] based on the findings from the identified studies. The evidence was then assessed using the modified grading of recommendations assessment, development and evaluation (GRADE) approach [31, 32]. The GRADE approach combines the rating and the ROB outcome into a final recommendation for each measurement property.

## Results

### Stage one

Figure 1 shows the PRISMA diagram for the stage one searches.

Our stage one searches initially yielded a large number of studies, however, after duplicate removal this number almost halved. A number of studies ($n = 38$) were excluded as they did not utilise PROMs, with most focussing on radiological outcome assessment and highlights the traditional focus in the literature on 'improving X-rays' which does not necessarily lead to an improvement in PROMs [33] or HR-QOL. A further 17 studies were excluded from the stage one searches as they focussed on AIS, and in particular those under 18 years of age, rather than AS.

The list of PROMs is included in the final box of Fig. 1 with the multiple uses of each PROM tallied across studies. The agreement between reviewers was 92.1% for the eligibility of studies at the stage one searches. After
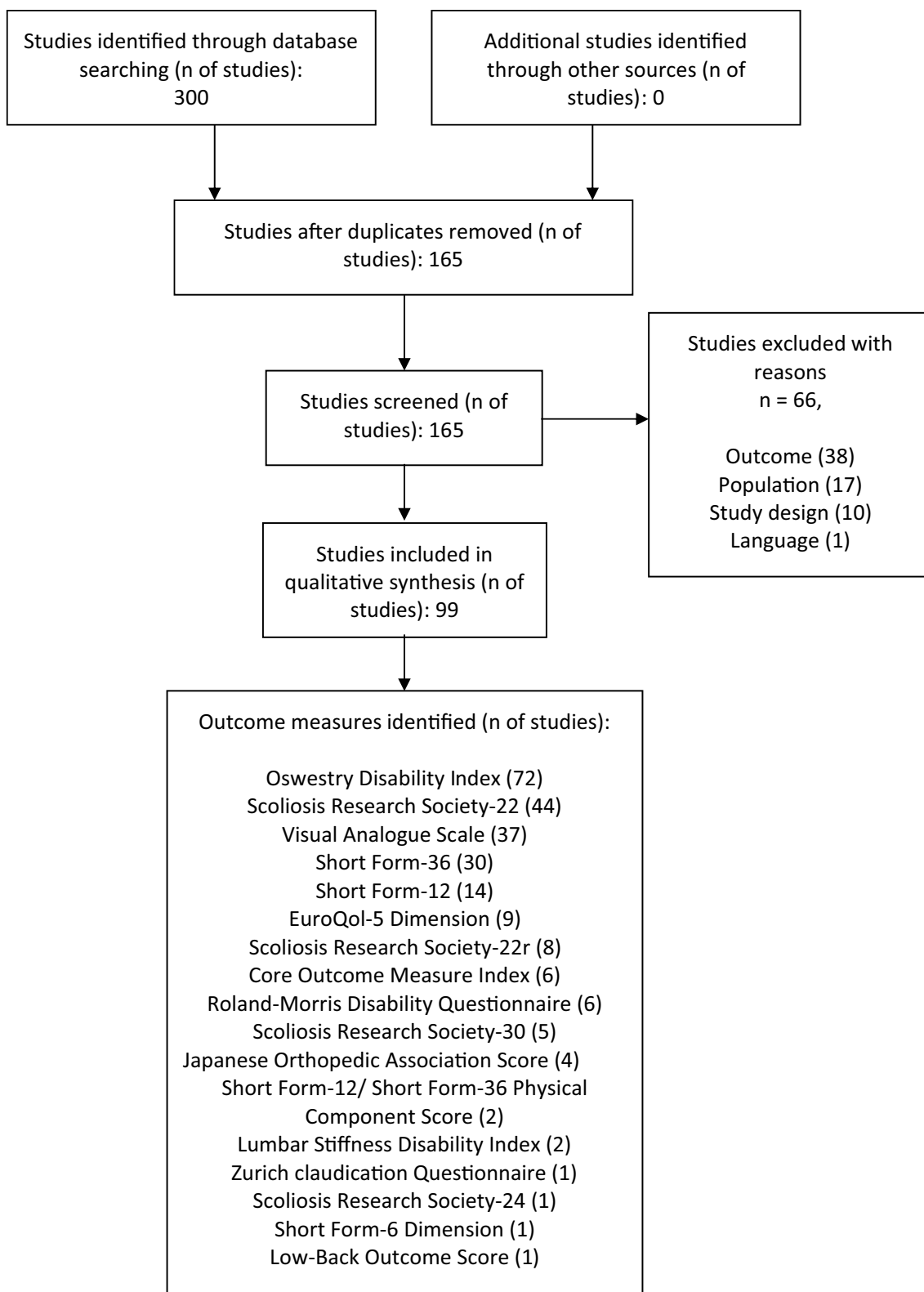
Studies identified through database searching (n of studies): 300

Additional studies identified through other sources (n of studies): 0

Studies after duplicates removed (n of studies): 165

Studies screened (n of studies): 165

Studies excluded with reasons
n = 66,

Outcome (38)
Population (17)
Study design (10)
Language (1)

Studies included in qualitative synthesis (n of studies): 99

Outcome measures identified (n of studies):

Oswestry Disability Index (72)
Scoliosis Research Society-22 (44)
Visual Analogue Scale (37)
Short Form-36 (30)
Short Form-12 (14)
EuroQol-5 Dimension (9)
Scoliosis Research Society-22r (8)
Core Outcome Measure Index (6)
Roland-Morris Disability Questionnaire (6)
Scoliosis Research Society-30 (5)
Japanese Orthopedic Association Score (4)
Short Form-12/ Short Form-36 Physical Component Score (2)
Lumbar Stiffness Disability Index (2)
Zurich claudication Questionnaire (1)
Scoliosis Research Society-24 (1)
Short Form-6 Dimension (1)
Low-Back Outcome Score (1)

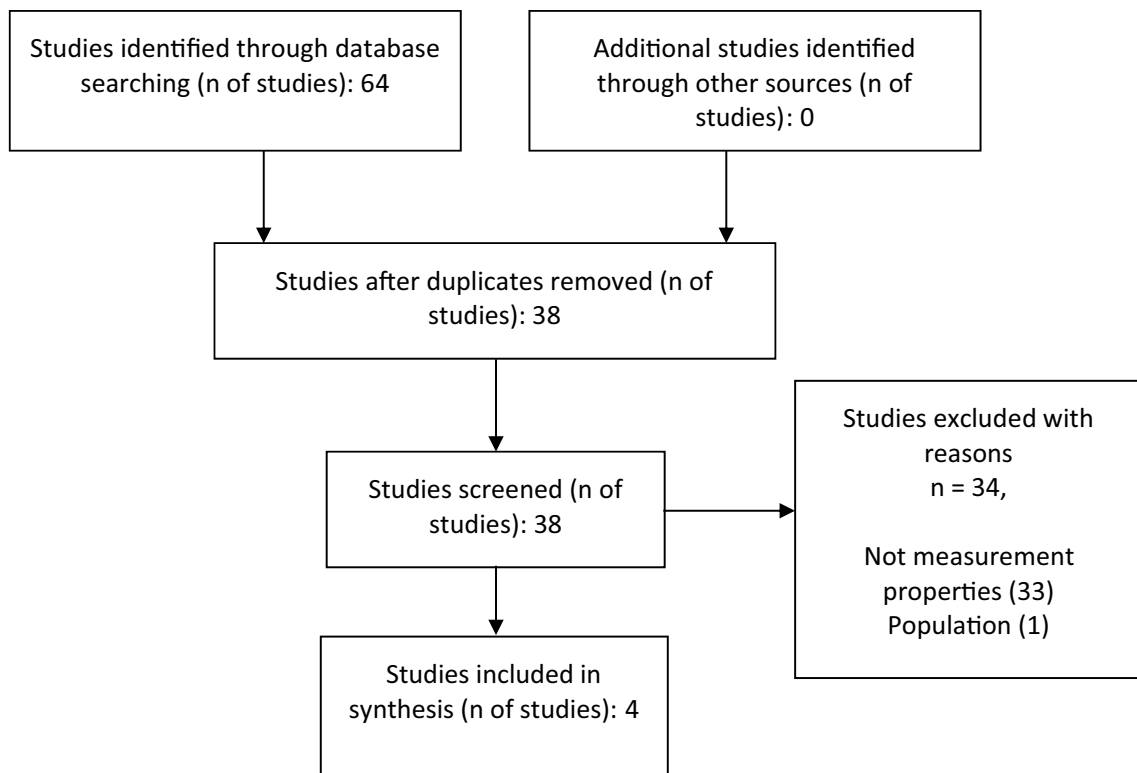**Fig. 1** PRISMA diagram for the stage one search

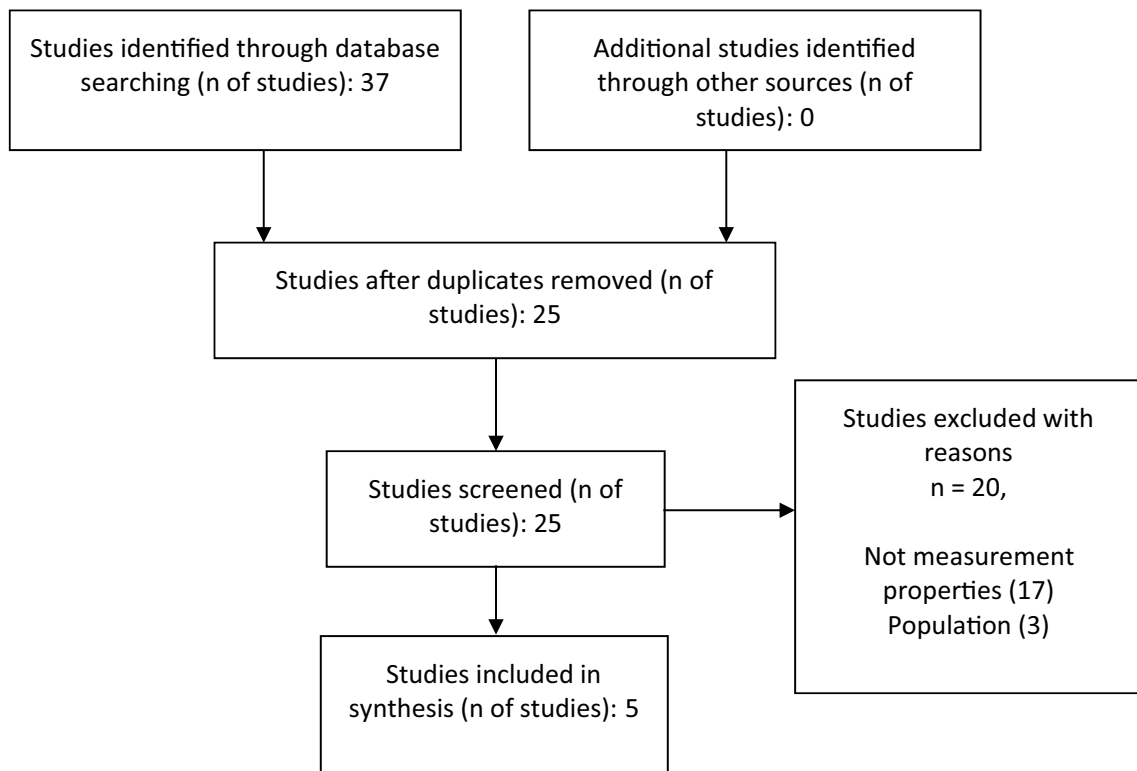**Fig. 2** PRISMA diagram for the stage two search for generic PROMs



**Fig. 3** PRISMA diagram for the stage two search for disease-specific PROMs

**Table 1** PROMs separated into generic and disease-specific groups

| Generic PROMs | Disease-specific PROMs |
| --- | --- |
| Oswestry Disability Index | SRS-22 |
| Visual Analogue Scale | SRS-22R |
| Short Form-36 | SRS-30 |
| Short Form-12 | SRS-24 |
| EuroQOL 5-Dimension | Core Outcome Measure Index |
| Roland-Morris Disability Questionnaire | Japanese Orthopaedic Association Score |
| Lumbar Stiffness Disability Index | |
| Zurich Claudication Score | |
| Short Form-6 Dimension | |
| Low Back Outcome Score | |

discussion, 100% agreement was achieved and the third reviewer was not consulted.

The references for the studies, grouped by PROM, found in search one are included as supplementary file 4.

### Stage two

Due to the large number of identified PROMs from stage one, the stage two search results were grouped to those which were generic PROMs and those which were disease-specific PROMs. The PROMs that were included in each section are shown in Table 1.

The stage two searches excluded most of the papers as they did not assess the measurement properties of the PROMs reported.

The PRISMA diagrams (Figs. 2 and 3) show the exclusion criteria which led to the selection of four studies that assessed the measurement properties of generic PROMs and five studies that assessed the measurement properties of disease-specific PROMs. Four studies [34–37] were results from both searches with one study [38] found in the disease-specific PROM searches alone. There was 100% agreement achieved between the two reviewers on screening and so a third reviewer was not consulted.

### Study characteristics

The stage one search identified 99 studies which utilised 16 different PROMs (Fig. 1). There were ten generic PROMs and six disease-specific PROMs (Table 1). Of the 16 PROMs, only three measures achieved the selection criteria of the stage two search. These were the SRS-22, SRS-22r and the ODI. As shown in Fig. 1, the ODI and SRS-22 were the most commonly used PROMs identified. Five studies were identified from the stage two searches and Table 2 shows the characteristics of these studies.

### SRS-22

The SRS-22 is a PROM that has been developed for the AIS population and has been through multiple revisions [19–22]. As shown in Table 3, this PROM covers 5 sub-domains, which were important to patients with AIS.

Responsiveness of the SRS-22 was studied by Bridwell et al., 2007 [34] who prospectively recruited 56 individuals to complete the SRS-22, ODI and SF-12 before and after surgery.

The validity and reliability of the SRS-22 was studied in individuals with AS by Berven et al. [35] who performed an observational study comparing healthy individuals to those with AS. The groups appear well matched for age and gender but there was no information on the curve types seen in the individuals with scoliosis.

The cross-cultural validity of the Russian version of SRS-22 was studied by Gubin et al. [36], the only study identified which was not performed in the USA. The study shows a discrepancy in the reported data concerning gender as it reports a total of 56 individuals comprising 25 males and 25 females. The authors were contacted for clarification, but no response was received.

### ODI

The ODI is a PROM for assessing low back pain and function in patients [39, 40]. It has been utilised in AS because of the importance of back pain in HR-QoL for these patients [10–12]. As shown in Table 3, its sub-domains focus on specific aspects of function which are relevant to adult patients.

The responsiveness of the ODI was the only property assessed in individuals with AS by Blizzard et al. [37]. Unfortunately, there is limited information provided in the published work as illustrated in Table 2.

**Table 2** A table of the identified studies and the characteristics of the participants from the stage two searches

| References | Country | Sample size | Age | Gender | Scoliosis curve type | PROM |
|---|---|---|---|---|---|---|
| Bridwell et al. [34] | USA | $n=56$ | 21–40 years ($n=16$)<br>41–60 years ($n=33$)<br>$>60$ years ($n=7$) | Female ($n=50$)<br>Male ($n=6$) | Double major curve ($n=15$)<br>Thoracic curves ($n=18$)<br>Thoracolumbar or lumbar curve ($n=22$)<br>Classified as other ($n=1$) | SRS-22 |
| Berven et al. [35] | USA | Individuals with scoliosis ($n=146$)<br>Individuals without scoliosis ($n=34$) | Average 47.1 years (range 18–88)<br>Average 46 years (range 19–71) | Female ($n=133$)<br>Male ($n=13$)<br>Female ($n=32$)<br>Male ($n=2$) | Not reported | SRS-22 |
| Gubin et al. [36] | Russia | Individuals with scoliosis ($n=140$)<br>Individuals without scoliosis ($n=56$) | Outpatients with scoliosis:<br>Average 57.5 years (range 23–75)<br>Inpatients with scoliosis:<br>Average 53.8 years (range 26–83)<br>Inpatients without scoliosis:<br>Average 43.3 years (range 26–66)<br>Outpatients without scoliosis:<br>average 36.4 years (range 22–60) | Female ($n=107$)<br>Male ($n=33$)<br>Female ($n=25$)<br>Male ($n=25$) | Idiopathic scoliosis ($n=16$)<br>Degenerative scoliosis ($n=124$) | SRS-22 |
| Blizzard et. al. [37] | USA | $n=1850$ | Not reported | Not reported | Surgical patients ($n=851$)<br>Non-surgical patients ($n=999$) | ODI |
| Arima et al. [38] | Japan and USA | $n=122$ | Mean age 61.9 years | Female ($n=106$)<br>Male ($n=16$) | Degenerative scoliosis ($n=45$)<br>Idiopathic scoliosis ($n=23$)<br>Kyphosis, deformity after fracture or iatrogenic scoliosis ($n=54$) | SRS-22r |

## SRS-22r

The SRS-22r is a development of the SRS-22 and involved a change to one of the questions in the function sub-domain, aimed at improving internal consistency in patients under the age of 18 [20].

Cross-cultural validity of the SRS-22r was the only measurement property investigated by Arima et al. [38]. Their study compared individuals in both the USA and Japan. The study did include a large number of individuals with other forms or spinal deformity ($n=54$), however, this was less than 50% and was, therefore, included.

## Measurement properties and narrative synthesis

Due to the small number of identified studies, a summary of the evidence was created, shown in Tables 4, 5 and 6. The tables are grouped by PROM to allow easily visualisation of which studies assessed which measurement properties.

Internal consistency was not assessed in line with the COSMIN guidance.

### SRS-22

The SRS-22 was the PROM that had the largest number of studies assessing the measurement properties, as can be seen in Table 4.

### Reliability

One study evaluated reliability of the SRS-22 [35]. Using the ROB scoring tool, the study was rated adequate for

**Table 3** Characteristics of the identified PROMs found in the stage two searches

| PROM | Country | Sub items | Target population | Mode of administration | Recall period | Response options | Translations available |
|---|---|---|---|---|---|---|---|
| SRS-22 | USA | Function<br>Pain<br>Self-image<br>Mental health<br>Satisfaction with management | Individuals with scoliosis | Self-administered | Now, post-surgery | 5 options | 17 languages [41–52] |
| ODI | UK | Pain intensity<br>Personal care<br>Lifting<br>Walking<br>Sitting<br>Standing<br>Sleeping<br>Sex life<br>Social life<br>Travelling | Adults with back pain | Self-administered | Now, post-surgery | 6 options | 10 languages [39, 53] |
| SRS-22r | USA | Function<br>Pain<br>Self-image<br>Mental health<br>Satisfaction with management | Individuals with scoliosis | Self-administered | Now, post-surgery | 5 options | 10 languages [45, 54–62] |

**Table 4** A description of the measurement properties of the SRS-22 ($n = 3$ studies)

| Measurement property | Studies | COSMIN Risk of Bias outcome | Overall rating | Quality of evidence (GRADE approach) |
|---|---|---|---|---|
| Reliability | Berven et al. [35] | Doubtful | Indeterminate | Very low |
| Content validity | Not assessed | Not assessed | Not assessed | Not assessed |
| Construct validity | Gubin et al. (CROSS-cultural validity) [36]<br>Berven et al. (hypothesis testing) [35] | Inadequate<br>Very good | Sufficient | Very low |
| Criterion validity | Not assessed | Not assessed | Not assessed | Not assessed |
| Responsiveness of the PROM | Bridwell et al. [34] | Inadequate | Sufficient | Very low |

**Table 5** A description of the measurement properties of the ODI ($n = 1$ studies)

| Measurement property | Studies | COSMIN Risk of Bias outcome | Overall rating | Quality of evidence (GRADE approach) |
|---|---|---|---|---|
| Reliability | Not assessed | Not assessed | Not assessed | Not assessed |
| Content validity | Not assessed | Not assessed | Not assessed | Not assessed |
| Construct validity | Not assessed | Not assessed | Not assessed | Not assessed |
| Criterion validity | Not assessed | Not assessed | Not assessed | Not assessed |
| Responsiveness of the PROM | Blizzard et al. [37] | Doubtful | Sufficient | Very low |

patient stability, which in this context relates to the individuals condition remaining stable between testing. It is assumed patients were stable in this study, but this was not directly stated. The time interval was appropriate, and it was assumed that the test conditions were similar. The overall ROB rating was downgraded to doubtful, due to a

Pearson correlation coefficient being utilised to analyse the test–retest results, rather than an intraclass correlation coefficient calculation.

There was very low-quality evidence indicating indeterminate reliability of the SRS-22 and this was downgraded

**Table 6** A description of the measurement properties of the SRS-22r ($n = 1$ studies)

| Measurement property | Studies | COSMIN Risk of Bias outcome | Overall rating | Quality of evidence (GRADE approach) |
|---|---|---|---|---|
| Reliability | Not assessed | Not assessed | Not assessed | Not assessed |
| Content validity | Not assessed | Not assessed | Not assessed | Not assessed |
| Construct validity | Arima et al. (cross-cultural validity) [38] | Doubtful | Indeterminate | Very low |
| Criterion validity | Not assessed | Not assessed | Not assessed | Not assessed |
| Responsiveness of the PROM | Not assessed | Not assessed | Not assessed | Not assessed |

twice due to the doubtful ROB rating and the lack of further studies.

## Content validity

No evidence was found for the content validity of the SRS-22.

## Construct validity

One study evaluated the cross-cultural validity of the SRS-22 [36]. The samples had similar characteristics except for the group variable and was rated very good. The approach to data analysis was appropriate and a Cronbach alpha was calculated, giving a rating of very good. However, the sample size was small and so this led to a rating of inadequate. Furthermore, the errors identified in the reported data were an important flaw and so, therefore, was again given an inadequate rating. This led to an overall ROB rating of inadequate.

One study evaluated the hypothesis testing of the SRS-22 [35]. The study was rated very good as the comparator instrument (SF-36) was clear and well defined. The measurement properties of the comparator are well-known and Pearson correlation coefficient and Cronbach alpha were utilised as the data analysis method leading to ratings of very good. There were no other flaws identified and, therefore, a ROB rating of very good.

There was very low-quality evidence indicating sufficient construct validity of the SRS-22. This was downgraded twice for ROB and once for imprecision.

## Criterion validity

No evidence was identified for criterion validity of the SRS-22.

## Responsiveness

The responsiveness of the SRS-22 was evaluated by one study [34]. The study was rated very good for the comparator instrument (ODI). However, it was rated inadequate for measurement properties of the comparator instrument as these were not described or discussed. The statistical method was also rated as inadequate as it is not clearly stated how the scores were converted to 'standard scores' and were then compared using paired $t$ tests. This led to an overall ROB rating of inadequate.

There was very low-quality evidence indicating sufficient responsiveness of the SRS-22. This was downgraded twice due to the ROB.

## ODI

One study evaluated the responsiveness of the ODI [37]. This study lacked an adequate description of the characteristics of the sub-groups leading to a rating of doubtful. The method of statistical analysis appears appropriate and was rated very good. However, the study is reported as an abstract and, therefore, very limited information is available. The ROB rating was doubtful.

There was very low-quality evidence indicating sufficient responsiveness of the ODI. The ROB rating led to 2 downgrades of the rating.

Table 5 demonstrates that no other studies were identified which assessed the other measurement properties of the ODI in AS.

## SRS-22r

One study assessed the cross-cultural validity of the SRS-22r [38]. It was unclear whether the samples were similar between the two groups which were compared, leading to a rating of doubtful. The statistical methodology used was receiver operating characteristic (ROC) curve analysis and appropriate statistical methods for the data and was rated very good. However, the sample size was also small and led to a rating of doubtful. The overall ROB rating was, therefore, doubtful.

There was very low-quality evidence indicating indeterminate cross-cultural validity for the SRS-22r. The rating was downgraded twice due to the ROB.

Table 6 demonstrates that no other studies were identified which assessed the remaining measurement properties of the SRS-22r in AS.

## Discussion

This is the first systematic review to synthesise HR-QOL PROMs for use in an AS population and evaluate their measurement properties. The objective was to identify the PROMs that are best suited to assessing HR-QOL in individuals with AS.

The stage one searches highlighted a wide array of PROMs being utilised, most notably generic PROMs. This may be explained by the fact that there is not a set of core outcome measures for AS, as identified by the COMET initiative [41] and defined for this population leading to ambiguity about the choice of PROMs to use. The most frequently utilised PROM is the ODI, which was designed as a tool for use in individuals with low back pain [39, 40], rather than individuals with AS. However, one of the most important reasons for the use of the ODI in this group is its role in assessing pain and function which are key considerations in HR-QOL [1]. The SRS-22 was the next most used PROM, and has been extensively utilised in individuals with AIS. Due to its development within the AIS population, the SRS-22 is highly likely to have a high internal consistency as it is a PROM that has been developed specifically for individuals with scoliosis. While the SRS-22 is not specifically designed for only adult individuals, this internal consistency would likely lead to a high degree of inter-relatedness between the items. However, it is also reasonable to assume that it will not assess all aspects of HR-QOL that are important for individuals with AS. These individuals represent a very different cohort to those with AIS and whether the same features that are assessed by the SRS-22 are as important to adults as they are adolescents [42] cannot be assumed.

Limited studies were identified investigating the ODI or SRS-22r, and the evidence that was identified was of very low quality. The ODI appears to have been widely adopted due to its relevance to adult populations and its focus on pain and function which are regarded to be more important features in adults. As back pain is one of the primary complaints of individuals with AS, the ODI could hold an important role as a PROM in this cohort. While the ODI may be an entirely reasonable PROM to use in this cohort, the lack of assessment of measurement properties mean it cannot currently be recommended. This lack of evidence should not be interpreted as a lack of evidence for the ODI as a PROM, but that there is insufficient evidence to support its use in individuals with AS.

The SRS-22 demonstrated very low-quality evidence to support its reliability, construct validity and responsiveness.

This result suggests that while SRS-22 is well studied in AIS, it has not had the measurement properties sufficiently assessed in AS to currently support its use. The lack of assessment of content validity is particularly important as it is often considered the most important measurement property of a PROM [43]. Without studies evaluating the content validity in individuals with AS, we may be missing domains that are highly valued by this group [43]. The SRS-22 aims to assess mental health, self-image and satisfaction with management, features absent from the ODI. However, the ODI appears to offer more granularity on function than the SRS-22 and this may explain its widespread use in AS.

Findings from this review therefore suggest that until the measurement properties of the ODI and SRS-22 are adequately assessed in AS, neither of these PROMs can be recommended.

### Strengths and limitations

This was a robust review assessing the measurement properties of PROMs for AS [25], following an a priori, published protocol to reduce bias. The review used a team who are experts at systematic review methodology and focussed on a very specific subject area. The two-stage search strategy allowed for the development of a comprehensive list of PROMs which have been used in this population and then allowed assessment of whether the measurement properties have been studied. The small number of studies identified by this study is a potential limitation.

### Implications for research and clinical practice

Further low risk of bias studies are needed to strengthen the level of evidence for the use of PROMs in AS, especially if these are to be used to ensure that interventions provide actual benefit to individuals' HR-QOL and ensure appropriate benefit relative to the risk and cost of treatments.

## Conclusion

Individuals with AS are being assessed using a wide array of PROMs. A very small number of studies have attempted to assess the measurement properties of these PROMs within an AS population. There is very low-quality evidence indicating indeterminate reliability, very low-quality evidence indicating sufficient construct validity and very low-quality evidence indicating sufficient responsiveness of the SRS-22. There is very low-quality evidence indicating sufficient responsiveness of the ODI. There is very low-quality evidence indicating indeterminate cross-cultural validity for the SRS-22r. Due to this low-quality evidence, no PROM can currently be recommended in this cohort and further studies

on the measurement properties of PROMs for individuals with AS are urgently required.

**Author contributions** JEA: Conceptualisation, design, literature review, analysis of data, drafting and final approval. CB: Conceptualisation, design, literature review, analysis of data, drafting and final approval. AG: Conceptualisation, design, literature review, analysis of data, drafting and final approval. ABR: Conceptualisation, design, analysis of data, drafting and final approval. NRH: Conceptualisation, design, analysis of data, drafting and final approval.

## Declarations

**Conflicts of interest** The authors have no competing or conflicting interests to declare.

**Ethical approval** No ethical approval was sought for this study as it does not involve patients.

## References

1. World Health Organisation (2012) World Health Organisation Quality of Life
2. World Health Organisation (2002) The international classification of functioning, disability and health
3. Aebi M (2005) The adult scoliosis. Eur Spine J 14:925–948. https://doi.org/10.1007/s00586-005-1053-9
4. Pellisé F, Vila-Casademunt A, Ferrer M et al (2015) Impact on health related quality of life of adult spinal deformity (ASD) compared with other chronic conditions. Eur Spine J 24:3–11. https://doi.org/10.1007/s00586-014-3542-1
5. Robin GC, Span Y, Steinberg R et al (1982) Scoliosis in the elderly: a follow-up study. Spine (Phila Pa 1976) 7:355–359. https://doi.org/10.1097/00007632-198207000-00005
6. Kostuik JP, Bentivoglio J (1981) The incidence of low-back pain in adult scoliosis. Spine (Phila Pa 1976) 6:268–273. https://doi.org/10.1097/00007632-198105000-00009
7. Pérennou D, Marcelli C, Hérisson C et al (1994) Adult lumbar scoliosis. Epidemiologic aspects in a low-back pain population. Spine (Phila Pa 1976) 19:123–128. https://doi.org/10.1097/00007632-199401001-00001
8. Jimbo S, Kobayashi T, Aono K et al (2012) Epidemiology of degenerative lumbar scoliosis: a community-based cohort study. Spine 37:1763–1770. https://doi.org/10.1097/BRS.0b013e3182575eaa
9. McAviney J, Roberts C, Sullivan B et al (2020) The prevalence of adult de novo scoliosis: a systematic review and meta-analysis. Eur Spine J 29:2960–2969. https://doi.org/10.1007/s00586-020-06453-0
10. Birknes JK, Harrop JS, White AP et al (2008) Adult degenerative scoliosis. Neurosurgery 63:A94–A103. https://doi.org/10.1227/01.NEU.0000325485.49323.B2
11. Palmisani M, Dema E, Cervellati S (2013) Surgical treatment of adult degenerative scoliosis. Eur Spine J 22:829–833. https://doi.org/10.1007/s00586-013-3012-1
12. McCormick JD, Werner BC, Shimer AL (2013) Patient-reported outcome measures in spine surgery. J Am Acad Orthop Surg 21:99–107. https://doi.org/10.5435/JAAOS-21-02-99
13. Bridwell KH, Shufflebarger HL, Lenke LG et al (2000) Parents' and patients' preferences and concerns in idiopathic adolescent scoliosis: a cross-sectional preoperative analysis. Spine 25:2392–2399. https://doi.org/10.1097/00007632-200009150-00020
14. Everett CR, Patel RK (2007) A systematic literature review of nonsurgical treatment in adult scoliosis. Spine 32:S130–S134. https://doi.org/10.1097/BRS.0b013e318134ea88
15. Wong E, Altaf F, Oh LJ et al (2017) Adult degenerative lumbar scoliosis. Orthopedics 40:e930–e939. https://doi.org/10.3928/01477447-20170606-02
16. Diebo BG, Shah NV, Boachie-Adjei O et al (2019) Adult spinal deformity. Lancet 394:160–172. https://doi.org/10.1016/S0140-6736(19)31125-0
17. Terran J, McHugh BJ, Fischer CR et al (2014) Surgical treatment for adult spinal deformity: projected cost effectiveness at 5-year follow-up. Ochsner J 14:14–22
18. Bagó J, Climent JM, Pérez-Grueso FJS et al (2013) Outcome instruments to assess scoliosis surgery. Eur Spine J 22:195–202. https://doi.org/10.1007/s00586-012-2352-6
19. Asher MA, Min Lai S, Burton DC (2000) Further development and validation of the scoliosis research society (SRS) outcomes instrument. Spine 25:2381–2386. https://doi.org/10.1097/00007632-200009150-00018
20. Asher MA, Lai SM, Glattes RC et al (2006) Refinement of the SRS-22 Health-Related Quality of Life Questionnaire Function Domain. Spine 31:593–597. https://doi.org/10.1097/01.brs.0000201331.50597.ea
21. Chen AF, Bi W, Singhabahu D et al (2013) Converting scoliosis research society-24 to scoliosis research society-22r in a surgical-range, medical/interventional adolescent idiopathic scoliosis patient cohort. Spine Deform 1:108–114. https://doi.org/10.1016/j.jspd.2012.12.003
22. Haher TR, Gorup JM, Shin TM et al (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. https://doi.org/10.1097/00007632-199907150-00008
23. Pynsent PB (2001) Choosing an outcome measure. J Bone Joint Surg Br 83:792–794. https://doi.org/10.1302/0301-620x.83b6.11973
24. Ioannidis JPA, Greenland S, Hlatky MA et al (2014) Increasing value and reducing waste in research design, conduct, and analysis. Lancet 383:166–175. https://doi.org/10.1016/S0140-6736(13)62227-8
25. Mokkink LB, Prinsen CAC, Patrick DL et al (2018) COSMIN methodology for systematic reviews of Patient-Reported outcome measures (PROMs). https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf
26. Archer JE, Baird C, Gardner A et al (2021) Evaluating measures of quality of life in adult scoliosis: a protocol for a systematic review and narrative synthesis. Syst Rev. https://doi.org/10.1186/s13643-021-01811-5
27. Page MJ, McKenzie JE, Bossuyt PM et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. https://doi.org/10.1136/bmj.n71
28. Cobb J (1948) Outline for the study of scoliosis. AAOS Instruct Course Lect 5:261–275
29. Dodd S, Clarke M, Becker L et al (2018) A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. J Clin Epidemiol 96:84–92. https://doi.org/10.1016/j.jclinepi.2017.12.020

30. PRISMA-P Group, Moher D, Shamseer L et al (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 4:1. https://doi.org/10.1186/2046-4053-4-1

31. Balshem H, Helfand M, Schünemann HJ et al (2011) GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 64:401–406. https://doi.org/10.1016/j.jclinepi.2010.07.015

32. Guyatt G, Oxman AD, Akl EA et al (2011) GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 64:383–394. https://doi.org/10.1016/j.jclinepi.2010.04.026

33. D'Andrea LP, Betz RR, Lenke LG et al (2000) Do radiographic parameters correlate with clinical outcomes in adolescent idiopathic scoliosis? Spine 25:1795–1802. https://doi.org/10.1097/00007632-200007150-00010

34. Bridwell KH, Berven S, Glassman S et al (2007) Is the SRS-22 instrument responsive to change in adult scoliosis patients having primary spinal deformity surgery? Spine 32:2220–2225. https://doi.org/10.1097/BRS.0b013e31814cf120

35. Berven S, Deviren V, Demir-Deviren S et al (2003) Studies in the modified scoliosis research society outcomes instrument in adults: validation, reliability, and discriminatory capacity. Spine 28:2164–2169. https://doi.org/10.1097/01.BRS.0000084666.53553.D6

36. Gubin A, Prudnikova O, Kamysheva V et al (2017) Clinical testing of the russian version of the SRS-22 questionnaire for adult scoliosis patients. Hir Pozvonoč 14:31–40. https://doi.org/10.1453/ss2017.2.31-40

37. Blizzard DJ, Sheets C, Isaacs RE et al (2016) Clinical outcomes of operative versus nonoperative treatment of adult scoliosis: a comparison using four standardized outcomes measures. Spine J 16:S309–S310. https://doi.org/10.1016/j.spinee.2016.07.234

38. Arima H, Carreon LY, Glassman SD et al (2019) Cultural variations in the minimum clinically important difference thresholds for SRS-22R after surgery for adult spinal deformity. Spine Deform 7:627–632. https://doi.org/10.1016/j.jspd.2018.10.003

39. Fairbank JC, Pynsent PB (2000) The Oswestry Disability Index. Spine (Phila Pa 1976) 25:2940–2952. https://doi.org/10.1097/00007632-200011150-00017

40. Fairbank JC, Couper J, Davies JB et al (1980) The Oswestry low back pain disability questionnaire. Physiotherapy 66:271–273

41. Williamson PR, Altman DG, Bagley H et al (2017) The COMET Handbook: version 1.0. Trials 18:280. https://doi.org/10.1186/s13063-017-1978-4

42. Mannion AF, Elfering A, Bago J et al (2018) Factor analysis of the SRS-22 outcome assessment instrument in patients with adult spinal deformity. Eur Spine J 27:685–699. https://doi.org/10.1007/s00586-017-5279-0

43. Terwee CB, Prinsen CAC, Chiarotto A et al (2018) COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 27:1159–1170. https://doi.org/10.1007/s11136-018-1829-0

## Authors and Affiliations

**James E. Archer[1]** · **Charles Baird[1]** · **Adrian Gardner[1,3]** · **Alison B. Rushton[2]** · **Nicola R. Heneghan[3]**

James E. Archer
jamesedward.archer@nhs.net

Charles Baird
charles.baird1@nhs.net

Alison B. Rushton
arushto3@uwo.ca

Nicola R. Heneghan
n.heneghan@bham.ac.uk

[1]  The Royal Orthopaedic Hospital NHS Foundation Trust, Bristol Road South, Northfield, Birmingham B31 2AP, UK

[2]  School of Physical Therapy, Faculty of Health Sciences, Western University, London, Canada

[3]  Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, UK