

## Minimal important differences of the SRS-22 Patient Questionnaire following surgical treatment of idiopathic scoliosis

Juan Bagó · Francisco J. S. Pérez-Grueso ·  
Esther Les · Pablo Hernández · Ferran Pellisé

Received: 13 January 2009 / Accepted: 1 June 2009 / Published online: 16 June 2009  
© Springer-Verlag 2009

**Abstract** The responsiveness of an instrument measuring health-related quality of life is an important indication of its construct validity. The SRS-22 Patient Questionnaire has become the most widely used patient-reported outcome instrument in the clinical evaluation of patients with idiopathic scoliosis. The responsiveness of the SRS-22 following surgical treatment in patients with idiopathic scoliosis has not been fully assessed. The aim of this study is to evaluate this factor by calculating the minimal important differences (MIDs) of the SRS-22 Questionnaire. The study included 91 patients with idiopathic scoliosis (77 females and 14 males), who underwent surgical treatment; mean age at the time of surgery was 18.1 years. Patients completed the SRS-22 questionnaire before surgery and at a follow-up visit (mean follow-up, 45.6 months). At follow-up, patients rated their overall situation as related to before surgery with a four-point Likert scale: 1—Worse, 2—Same, 3—Better, 4—Much Better. This evaluation represented the global perceived effect (GPE) and served as the anchor criterion for calculating the MID. MIDs were calculated using two approaches. The anchor-based MID (MID-A) was defined as the mean preoperative/follow-up difference in SRS-22 scores in the group of patients who stated they were much better than before surgery (GPE = 4). Using the same anchor criterion, the optimal

cut-off value able to identify patients that had clearly improved was determined on a receiver operating characteristic (ROC) curve. In addition, the distribution-based MID (MID-D) was calculated by the standard error of measurement method. The MID-As found for the different subscales and the sum score were: pain 0.6, function 0.3, image 1.3, mental health 0.3, average sum score 0.6, and raw sum score 13.1. The cut-off values on the ROC curve were: pain 0.2, function 0.0, image 1.6, mental health 0.4, average sum score 0.4, and raw sum score 10. The MID-Ds were: pain 0.6, function 0.8, image 0.5, mental health 0.4, average sum score 0.5, and raw sum score 6.8. As was expected, the MID values differed according to the calculation method used. In light of the fact that the MID-As for the function and mental health subscales are below the measurement error of the instrument, it seems preferable to use the MID-D values for determining subscale changes. If the purpose is to analyze sum score changes (either the raw or average values), the MID-A is preferable because it includes the patient's evaluation of the results of surgery.

**Keywords** Idiopathic scoliosis · Health-related-quality-of-life · SRS-22 Patient Questionnaire · Minimal important difference

J. Bagó (✉) · E. Les · F. Pellisé  
Department of Orthopaedic Surgery,  
Hospital Universitario Vall d'Hebron,  
P<sup>o</sup> Vall d'Hebron, 119, 08035 Barcelona, Spain  
e-mail: jbagó@vhebron.net; 13811jbg@comb.es

F. J. S. Pérez-Grueso · P. Hernández  
Department of Orthopaedic Surgery,  
Hospital Universitario La Paz, P<sup>o</sup> de la Castellana, 261,  
28046 Madrid, Spain

### Introduction

The SRS-22 Patient Questionnaire has become the most widely used patient-reported outcome (PRO) instrument for evaluating individuals with idiopathic scoliosis. It has been properly validated in both adolescents [2, 4] and adults [10], and adaptations in several languages are now available [1, 6, 12, 13, 19, 22]. To determine the validity of the SRS-22, the internal consistency, test-retest reliability,

factorial analysis, floor and ceiling effects, and convergent-discriminant validity have been tested [4, 10, 14].

The responsiveness is another important factor related to the construct validity of the instrument. The simplest way to assess this characteristic is by determining whether clinically relevant changes are associated with statistically significant differences in the scores of the scale over time. Along this line, Asher et al. [3] analyzed the responsiveness of the SRS-22 questionnaire in 58 patients who underwent surgery for idiopathic scoliosis. Patients were assessed at 3, 6, 12 (only 38 patients) and 24 (only 19 patients) months. Statistically significant differences were found in the pain, function, and image subscales, and in the total score. There were no significant changes in the mental health subscale. Bridwell et al. [9] analyzed a series of 56 adult patients 2 years after the surgical intervention. These authors found a significant change in all the SRS-22 subscales and in the total score. In addition, the change in the SRS-22 Questionnaire was greater than that found for the Oswestry Disability Index and the scores of the SF-12 components.

Nonetheless, to interpret the effect of a treatment, it is not only necessary to determine whether there is a statistically significant difference; but also important to assess how relevant the effects are for the patients [16, 26]. For this reason, it is preferable to evaluate an instrument's responsiveness by determining the relationship between the score changes and the patient's self-reported clinical changes over time [23]. This approach leads to determination of the minimal important difference (MID), that is, the smallest difference in the score of the outcome instrument that informed patients perceive as important [11, 15, 16].

Currently, there is no consensus as to the best method for determining the MID. For a specific PRO, it is recommended to obtain a range of MIDs obtained by different methods [23]. Most commonly, two approaches are used for this purpose: anchor-based methods and distribution-based methods. Anchor-based methods compare the change in the PRO being examined with another measure of change. Most commonly, patients are requested to rate the global perceived effect (GPE) of the intervention. From these results, the MID is usually calculated as the mean change of the scores in patients that have clearly improved. This MID is also termed *minimal clinically important difference* [23]. Another method within the anchor-based group of methods is to determine the MID that best differentiates between patients that have clearly improved and the remaining patients. To this end, the optimal cut-off point is calculated on a receiver operating characteristic (ROC) curve.

The distribution-based methods are derived from the concept that the MID can be estimated from the

distribution of the scores [23]. One commonly used method is calculation of the standard error of measurement (SEM), that is, the measurement error inherent to the instrument. When calculating the SEM, an estimator of reliability is included (e.g., the intraclass correlation coefficient, ICC) [20]. This type of MID is also known as the *minimal detectable change* [25].

The aim of this study is to calculate the MID of the SRS-22 Questionnaire with the use of two anchor-based methods (mean score change and optimal cut-off value) using the GPE as the external criterion for change, and with a distribution-based method, calculation of the SEM.

## Methods

### *Study population*

Patients were enrolled in the two participating centers. The eligibility criteria were the following: a diagnosis of idiopathic scoliosis and scheduled surgical treatment to correct the deformity, age 10 to 40 years, a suitable radiologic study, and completion of the SRS-22 questionnaire. Between July 2001 and July 2006, 97 patients from the two centers were included in the study. After an interval of at least 2 years since surgery, patients were contacted for a radiological and clinical review for the study.

### Measurement instruments

#### *SRS-22 Patient Questionnaire*

The SRS-22 contains 22 questions covering 5 domains: function/activity 5 items; pain 5 items; self-perceived image 5 items; mental health 5 items; and satisfaction with treatment 2 items. Each item is scored from 1 (worst) to 5 (best). Each domain has a total sum score ranging from 5 to 25, except for *satisfaction*, which ranges from 2 to 10. The sum of the first 4 domains gives a maximum subtotal of 100, and when the satisfaction domain is included, the maximum total is 110. In the present paper, results are expressed as the mean (total sum of the domain divided by the number of items answered) for each domain and the subtotal score. Because the aim of this study is to analyze the MIDs of the different subscales, data on satisfaction had to be excluded, as most patients did not answer the questions about satisfaction before surgery. Thus, for the present study, the terms *sum score* and *total score* refer to the scores for the group of four scales, excluding the satisfaction scale. The total score was presented in three ways: the raw sum score without satisfaction (possible range 20–100), the average sum score (raw score/20), and the raw score percentage of improvement. The questionnaire used

was the revised version, which includes the modification of question 18 [5, 7].

### Global perceived effect (GPE)

The assessment of GPE was performed at the follow-up visit. Patients were asked to rate their overall situation in relation to before surgery with a 4-point Likert scale: 1, Worse; 2, Same; 3, Better; or 4, Much Better.

### Data analysis

MIDs were determined using three different methods:

1. Anchor-based method (MID-A): the MID was calculated as the mean of the preoperative/follow-up difference in the group of patients who stated they were much better than before surgery ( $GPE = 4$ ).
2. Optimal cut-off point or ROC curve analysis. The rationale of this method is the use of the questionnaire as an instrument to distinguish between patients who have improved and those who have not improved. To this end, the receiver operating characteristic (ROC) curve is determined. The anchor variable is the GPE, which functions as a reference standard that can differentiate between patients showing an important clinical change or not. The diagnostic precision of the measuring instrument (in this case the SRS-22 questionnaire) can be expressed in terms of sensitivity and specificity and be represented on an ROC curve. The optimal cut-off point is the score that provides the best sensitivity and specificity. In this case, the sensitivity is the number of patients correctly identified as having improved, whereas the specificity is the number correctly identified as not having improved. False-positive cases are patients who have not had an important clinical change according to their GPE, but whose score change is higher than the instrument's cut-off point. False negatives are individuals who consider that they are "much better" according to their GPE, but whose score change is below the instrument's cut-off point.
3. Distribution-based method (MID-D) or standard error of measurement (SEM) method: calculated for the entire sample ( $n = 91$ ) with the formula  $SEM = SD\sqrt{1 - \alpha}$ , in which SD is the standard deviation of the pretreatment score and  $\alpha$  is the reliability coefficient of the instrument. In this case, the intraclass correlation coefficient obtained in the test-retest reliability analysis was used [6]. Once the SEM was known, a formula was applied to estimate (at 95% confidence) that the change in the score exceeded the systematic measurement error. This change is:

$\Delta = 1.96\sqrt{2} \times SEM$ . The result is at 95% confidence, which is why 1.96 is used [8].

## Results

### Characteristics of the patients

Of the 97 patients included in the study, 4 were lost to follow-up and 2 refused to participate in the follow-up visit. Therefore, the analysis is based on 91 patients (93.8% response rate) (77 females and 14 males) with a mean age at the time of surgery of 18.1 years (range 10–38 years). The mean follow-up time was 45.6 months (range 24–87 months). All patients were diagnosed with idiopathic scoliosis and were scheduled to receive surgical treatment. The curve pattern was determined according to the Lenke classification [21]: 32 cases were type 1, 17 type 2, 10 type 3, 5 type 4, 10 type 5, and 17 type 6. The mean magnitude of the upper thoracic curve was  $49.8^\circ$ , the main thoracic curve  $61.5^\circ$ , and the thoracolumbar/lumbar curve  $60.8^\circ$ . The type of surgery performed included posterior spinal fusion and instrumentation in 74 cases, anterior spinal fusion and instrumentation in 8 cases, and anterior and posterior spinal fusion and instrumentation in 9 cases. At the time of the follow-up visit, the average magnitude of the upper thoracic, main thoracic and thoracolumbar/lumbar curve was  $32.8^\circ$ ,  $29.4^\circ$ , and  $30.1^\circ$ , respectively.

### Outcome measures

The scores for the SRS-22 scales before surgery and at follow-up are shown in Table 1. A statistically significant improvement was seen in the sum score and in the pain, image, and mental health scales, whereas the improvement in the function scale was not significant. The mean percentage improvement of the raw sum score was 13.9% (range  $-27.6$  to  $75.4\%$ ).

In the follow-up interview, 47 patients (51.6%) considered they were much better than before surgery ( $GPE = 4$ ),

**Table 1** Mean scores for the SRS-22 Patient Questionnaire scales before surgery and at follow-up

	Preop	Follow-up	Mean change
Pain	3.9	4.2	0.3*
Function	4.2	4.3	0.1
Image	2.9	3.9	1.0**
Mental health	3.8	3.9	0.1*
Raw sum	74.1	82.5	8.4**
Average sum	3.7	4.1	0.4**

\*  $P \leq 0.05$ , \*\*  $P < 0.0001$

37 considered they were better (GPE = 3), 5 stated they were the same (GPE = 2), and 2 stated they were worse (GPE = 1). The validity of the scale assessing the global perceived effect was supported by the correlation between the GPE ratings and the preoperative/follow-up difference in the SRS-22 raw sum score ( $r = 0.4, P = 0.0001$ ).

**Minimal important differences**

The mean SRS-22 scale scores and sum score (raw and average) for the group with an evident improvement (GPE = 4) and for the remaining patients (GPE < 4) are shown in Table 2. In addition, the percentage of improvement of the raw sum score is reported for each group. The mean score changes for the group with GPE = 4 was significantly greater ( $t$  test  $P < 0.05$ ) for all the subscales and the sum score than the changes observed in the group with GPE < 4. The mean score change of the GPE = 4 group represents the minimal important difference using GPE as the anchor criterion (MID-A). The SEM was calculated using the system described in “Methods”, and the minimal importance difference was determined with the distribution-based method (MID-D).

The MID-A and MID-D, as well as the optimal cut-off points of the ROC curves, with the respective sensitivity, specificity, and area-under-the-curve (AUC) are shown in Table 3. The AUC represents the probability that the cut-off correctly differentiates between the two groups of patients (GPE = 4 vs. GPE < 4). In the way of orientation, an AUC of 0.7–0.8 is considered acceptable, whereas an AUC of 0.8–0.9 is considered excellent [15].

**Discussion**

This study provides data on the various MIDs for the subscales and sum score of the SRS-22 patient questionnaire in a group of patients who underwent surgery for

**Table 3** Minimal important difference according to anchor-based and distribution-based methods and optimal cut-off scores of the ROC curves

	MID-A	MID-D	Cut-off	Sensitivity	Specificity	AUC
Pain	0.6	0.6	0.2	72.3	63.6	0.69
Function	0.3	0.8	0.0	74.4	52.2	0.65
Image	1.3	0.5	1.6	46.8	84	0.69
Mental health	0.3	0.4	0.4	51	65.9	0.61
Raw-Sum	13.1	6.8	10	59.7	84	0.71
Av-Sum	0.6	0.5	0.4	84	59	0.71
Improvement (%)	20.6		13.2	81.8	53.2	0.69

Data on sensitivity, specificity, and area under the curve (AUC) are for the cut-off points

idiopathic scoliosis. The sample is heterogeneous for age, with both adolescents and adults included. Nevertheless, the version of the questionnaire used has shown similar validity for the age range (10–40 years) included in the analysis [5, 7].

As was expected, the MID values differed substantially depending on the calculation method used. This has also been observed in other outcome instruments commonly applied in the evaluation of low back pain [17, 20, 25]. There is currently no consensus regarding the recommended methods for determining the MID [23].

Calculation by an anchor-based method is simple and the result is easy to interpret. The MID-A (or minimal clinically important difference) represents the change in the scale score of a group of patients selected according to their response on the GPE rating. Nonetheless, some researchers question the use of global ratings. One important criticism is that the validity of the GPE scale is unknown; nonetheless, in our case, several data support the scale’s validity. First, the changes in the various subscales showed a significant difference between the two groups of patients

**Table 2** Mean values of the subscales and sum score of the SRS-22 according to the GPE rating

	GPE = 4			GPE < 4		
	Preop	Fup	Mean change (95% CI)	Preop	Fup	Mean change
Pain	3.8	4.5	0.6 (0.3–0.8)	3.9	3.9	0.0 (–0.2–0.3)
Function	4.1	4.5	0.3 (0.1–0.5)	4.2	4.1	0.0 (–0.2–0.1)
Image	2.9	4.2	1.3 (1.1–1.6)	2.8	3.6	0.7 (0.4–0.9)
Mental health	3.8	4.2	0.3 (0.1–0.5)	3.7	3.7	0.0 (–0.2–0.2)
Raw-sum	74.1	87.2	13.1 (9.6–16.5)	74	77.4	3.4 (0.0–6.9)
Av-sum	3.7	4.3	0.6 (0.4–0.8)	3.7	3.8	0.1 (0.0–0.3)
% Improvement Raw-Score			20.6 (14.5–26.6)			6.8 (1.2–12.3)

Patients with a GPE rating other than “much better” are grouped as GPE < 4

Mean change in the GPE = 4 group represents the MID using the anchor-based method and is presented with the 95% confidence interval

(GPE = 4 vs. GPE < 4). Second, there was a statistically significant correlation between the GPE rating and the score changes. This correlation is generally considered necessary [11, 15, 16, 23] to accept the validity of the rating score. Moreover, because of this association, the GPE acquires the aspect of a fast, valid tool for evaluating the patient in daily practice. This has also been seen in other PROs used to assess low back pain [18]. To assess the GPE, we used a scale including only four possible answers: much better, better, same, and worse. There is no consensus regarding how many categories a GPE rating scale should have. What is clear, however, is that the choice of patient group used as the external criterion for calculating the MID is arbitrary. The greater the number of levels, the smaller the difference between the adjacent levels, and the smaller will be the MID; hence, there is a risk that it will not exceed the inherent error of the measuring method. Although some experts have recommended the use of scales with seven levels [16], we opted for a scale with four levels, as has been applied in other studies investigating the outcome of spinal surgery [17].

Calculation of the MID-A in the present study was based on the score change in patients who considered their status “much better” than before surgery. This group was chosen after analyzing the SRS-22 score changes in each of the groups defined by the GPE. Taking the raw sum score as an example, we found that the mean change was 13 in the “much better” group, 4.8 in the “better” group,  $-0.8$  in the “same” group, and  $-12$  in the “worse” group. Thus, it is evident that patients in the “much better” group perceived an obvious improvement in their condition after the intervention, those in the “better” group only a slight improvement and those in the other groups believed they were not better. Thus, the MID-A represents a change from the patient’s viewpoint, that is what PROs attempt to express. Despite the potential problems with this method (recall bias, concern about anchor validity, etc.) it is the one recommended by most experts [23, 24].

The optimal cut-off point is another anchor-based method. Based on an external anchor criterion (GPE category in this case), analysis of the ROC curve determines a value that allows differentiation between patients who have improved and those who have not. In addition, data are obtained on the accuracy (area under the curve), sensitivity, and specificity of this cut-off value.

The MID-D (or minimal detectable change), described as the minimal amount of change that is not attributable to the noise or imprecision of the measurement instrument, is a more conservative method that is particularly useful in questionnaires such as the SRS-22, in which the difference between the various responses is subtle [25]. One advantage of the MID-D is that it is easier to generate than anchor-based data. The biggest criticism to this method is

that it is based purely on statistical calculations and does not take into account information provided by the patients. The MID-D, based on calculation of the SEM, represents the minimal score change, above which (with 95% certainty) it is guaranteed that the change is not due to measurement error. It provides useful information for assessing MID values obtained with anchor-based methods because it enables assessment of the distance between the MID and the measurement error [23].

We found that the MID-As of the function and mental health subscales were below the MID-D, whereas the MID-As of the image and pain subscales, as well as the sum score (both raw and average) were greater than the respective MID-Ds. In clinical research, when the MID-A value is lower than the MID-D, the validity of the instrument may be questioned because it cannot be guaranteed that the change observed is not due to an measurement error of the instrument. In contrast, when the MID-A is greater than the MID-D, the MID-A is the value of choice because in all probability it is a true change, perceived by both the patient and the instrument.

Using the optimal cut-off point method, we obtained lower MIDs (except in the image subscale) that were below the level of the measurement error. Moreover, the sensitivity and specificity were low, the curve did not have a smooth profile, and the area under the curve was far from excellent. We also calculated the MID-A of the percentage of improvement of the raw sum score (range, 13.2–20.6%). It has been suggested that the percentage of improvement can be useful in instruments (such as the SRS-22) that present a marked ceiling effect, that is, those in which patients with high baseline scores have little possibility to improve [15]. The analysis of the MIDs was only performed in relation to improvement. It was considered unfeasible to perform the analysis of change for relevant worsening because of the small number of patients involved (only two considered they were worse after surgery).

Because there is no consensus as to the optimal method for calculating the MID for a specific PRO instrument in a specific patient population, it is inevitable that a range of MID values must be managed. In the case of the SRS-22, the situation is complex because the questionnaire is divided into five scales (sum score and four subscales) each with its respective MID. The investigator must decide which MID is the most suitable, in keeping with the objectives of the clinical research. An MID value can be used to dichotomize a group of patients under study. Dichotomization facilitates comprehension of the results for clinicians because they are often unaware of the clinical significance of a specific raw score or score change [24]. Dichotomization of patients according to an MID value (e.g., the difference of mean values or a percentage of

patients exceeding this score) is useful for designing clinical studies (sample size calculation) and interpreting the results, and is the method recommended by regulating agencies, such as the Food and Drug Administration [11].

To assess the results of a specific treatment, the criteria for success/failure should be defined a priori, and it is crucial to know the MIDs of the measurement instrument for this task. Although it is not within the aim of the present study, the data presented allow a preliminary analysis of the effects of surgery on the clinical status of patients assessed with the SRS-22. The results show that patients have a clinically significant improvement only in pain, self-perceived body image, and overall quality of life. These data should be related to the characteristics of the study group. The study patients had moderately severe scoliosis (Cobb angle around 60°), in which the main clinical problem is the trunk deformity. Our data confirm a generalized clinical impression: in this range of curve magnitude, the impact of scoliosis is produced in the areas of body image and pain, and this has repercussions on the patient's general quality of life. It is logical that in this group of patients no significant changes are produced in function or mental health. Thus, the lack of responsiveness of the SRS-22 in these areas may correspond to a clinical reality: at baseline, before surgery, these patients have high scores in these domains and there is little margin for improvement. It would be interesting to assess the responsiveness of the instrument in groups of patients with more severe baseline status and lower scores on the subscales before surgery.

In the case of the SRS-22 and in accordance with our results, use of the MID-D should be recommended when the aim is to analyze each of the scales of the instrument (Pain 0.6; Function 0.8; Image 0.5; Mental Health 0.4; Raw-Sum 6.8; Average-Sum 0.5) because some subscales show an MID-A below the measurement error. Since the MID-D does not take into account the clinical importance of the change, it should be accompanied by data on the patients' satisfaction with the treatment [25]. This is not a problem because the SRS-22 has a valid and reliable scale on satisfaction [4]. However, if the interest is focused on the patient's response, we suggest using only the MID-A of the sum score (Raw-sum 13.1; Average-sum 0.6), which is higher than the MID-D. Thus, the SRS-22 would be presented with a single value, as occurs with other PRO instruments, such as the Oswestry Disability Index and the Roland–Morris Disability Questionnaire. The cut-off points obtained from the ROC curve seem to have little use in the SRS-22 because they are below the measurement error, and the sensitivity and specificity are low.

To our knowledge, this is the first study that has provided data on the MIDs of the SRS-22 Patient Questionnaire. In all probability, future studies in this line will yield

values somewhat different from the MIDs reported herein. As Revicki et al. [23] have suggested for other PRO instruments, it is likely that a single MID value will not be established for the SRS-22 until a systematic review involving several studies is available.

**Acknowledgments** The authors want to acknowledge Celine Cavallo for translating and editing the paper.

## References

- Alanay A, Cil A, Berk H, Acaroglu RE, Yazici M, Akcali O et al (2005) Reliability and validity of adapted Turkish Version of Scoliosis Research Society-22 (SRS-22) questionnaire. *Spine* 30:2464–2468. doi:[10.1097/01.brs.0000184366.71761.84](https://doi.org/10.1097/01.brs.0000184366.71761.84)
- Asher MA, Lai SM, Burton D, Manna B (2003) Discrimination validity of the Scoliosis Research Society-22 patient questionnaire: relationship to idiopathic scoliosis curve pattern and curve size. *Spine* 28:74–78. doi:[10.1097/00007632-200301010-00017](https://doi.org/10.1097/00007632-200301010-00017)
- Asher MA, Lai SM, Burton D, Manna B (2003) Scoliosis Research Society-22 patient questionnaire: responsiveness to change associated with surgical treatment. *Spine* 28:70–73. doi:[10.1097/00007632-200301010-00016](https://doi.org/10.1097/00007632-200301010-00016)
- Asher MA, Lai SM, Burton D, Manna B (2003) The reliability and concurrent validity of the Scoliosis Research Society-22 patient questionnaire for idiopathic scoliosis. *Spine* 28:63–69. doi:[10.1097/00007632-200301010-00015](https://doi.org/10.1097/00007632-200301010-00015)
- Asher MA, Lai SM, Glattes RC, Burton D, Alanay A, Bago (2006) Refinement of the SRS-22 health-related quality of life questionnaire function domain. *Spine* 31:593–597. doi:[10.1097/01.brs.0000201331.50597.ea](https://doi.org/10.1097/01.brs.0000201331.50597.ea)
- 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* 29:1676–1680. doi:[10.1097/01.BRS.0000132306.53942.10](https://doi.org/10.1097/01.BRS.0000132306.53942.10)
- Bago J, Climent JM, Ey A, Perez-Grueso FJS, Izquierdo E (2006) Letter to the Editor. Re: Asher Ma; Lai, SM; Glattes CR; Burton DC; Alanay A; Bago J.—Refinement of the SRS-22 health-related quality of life questionnaire function domain. *Spine*, 2006, 31(5): 593–597. *Spine* 31:1758. doi:[10.1097/01.brs.0000224188.59590.0e](https://doi.org/10.1097/01.brs.0000224188.59590.0e)
- Beaton DE (2000) Understanding the relevance of measured change through studies of responsiveness. *Spine* 25:3192–3199. doi:[10.1097/00007632-200012150-00015](https://doi.org/10.1097/00007632-200012150-00015)
- Bridwell KH, Berven S, Glassman S, Hamill C, Horton WC, Lenke LG et al (2007) Is the SRS-22 instrument responsive to change in adult scoliosis patients having primary spinal deformity surgery? *Spine* 32:2220–2225. doi:[10.1097/BRS.0b013e3181453e22](https://doi.org/10.1097/BRS.0b013e3181453e22)
- Bridwell KH, Cats-Baril W, Harrast J, Berven S, Glassman S, Farcy JP et al (2005) The validity of the SRS-22 instrument in an adult spinal deformity population compared with the Oswestry and SF-12: a study of response distribution, concurrent validity, internal consistency and reliability. *Spine* 30:455–461. doi:[10.1097/01.brs.0000153393.82368.6b](https://doi.org/10.1097/01.brs.0000153393.82368.6b)
- Brozek JL, Guyatt GH, Schunemann HJ (2006) How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health Qual Life Outcomes* 4:69. doi:[10.1186/1477-7525-4-69](https://doi.org/10.1186/1477-7525-4-69)
- Bunge EM, Juttman RE, de Kleuver M, van Biezen FC, de Koning HJ, NESICIO group (2007) Health-related quality of life in patients with adolescent idiopathic scoliosis after treatment:

- short-term effects after brace or surgical treatment. *Eur Spine J* 16:83–89. doi:[10.1007/s00586-006-0097-9](https://doi.org/10.1007/s00586-006-0097-9)
13. Cheung KM, Senkoylu A, Alanay A, Genc Y, Lau S, Luk KD (2007) Reliability and concurrent validity of the adapted Chinese version of Scoliosis Research Society-22 (SRS-22) questionnaire. *Spine* 32:1141–1145. doi:[10.1097/01.brs.0000261562.48888.e3](https://doi.org/10.1097/01.brs.0000261562.48888.e3)
  14. Climent JM, Bago J, Ey A, Perez-Gruoso FJ, Izquierdo E (2005) Validity of the Spanish version of the Scoliosis Research Society-22 (SRS-22). Patient Questionnaire. *Spine* 30:705–709
  15. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC (2007) Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 7:541–546. doi:[10.1016/j.spinee.2007.01.008](https://doi.org/10.1016/j.spinee.2007.01.008)
  16. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting Group (2002) Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 77:371–383
  17. Hägg O, Fritzell P, Nordwall A (2003) The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 12:12–20
  18. Hägg O, Fritzell P, Oden A, Nordwall A, Swedish Lumbar Spine Study Group (2002) Simplifying outcome measurement. Evaluation of instruments for measuring outcome after fusion surgery for low back pain. *Spine* 27:1213–1222. doi:[10.1097/00007632-200206010-00014](https://doi.org/10.1097/00007632-200206010-00014)
  19. 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* 32:E141–E146. doi:[10.1097/01.brs.0000255220.47077.33](https://doi.org/10.1097/01.brs.0000255220.47077.33)
  20. Jordan K, Dunn KM, Lewis M, Croft P (2006) A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. *J Clin Epidemiol* 59:45–52. doi:[10.1016/j.jclinepi.2005.03.018](https://doi.org/10.1016/j.jclinepi.2005.03.018)
  21. Lenke LG, Betz RR, Harms J, Bridwell kh, Clements DH, Lowe TG et al (2001) Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis. *J Bone Joint Surg* 83-A:1169–1181
  22. Monticone M, Carabolona R, Negrini S (2004) Reliability of the Scoliosis Research Society-22 Patient Questionnaire (Italian version) in mild adolescent vertebral deformities. *Eura Medico-phys* 40:191–197
  23. Revicki D, Hays RD, Cella D, Sloan J (2008) Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 61:102–109. doi:[10.1016/j.jclinepi.2007.03.012](https://doi.org/10.1016/j.jclinepi.2007.03.012)
  24. Schünemann HJ, Akl EA, Guyatt GH (2006) Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. *Health Qual Life Outcomes* 4:62. doi:[10.1186/1477-7525-4-62](https://doi.org/10.1186/1477-7525-4-62)
  25. Van del Roer N, Ostelo RWJG, Bekkering GE, Van Tulder MW, De Vet HCW (2006) Minimal clinically important change for pain intensity, functional status, and general health status in patients with non-specific low back pain. *Spine* 31:578–582. doi:[10.1097/01.brs.0000201293.57439.47](https://doi.org/10.1097/01.brs.0000201293.57439.47)
  26. van Tulder M, Malmivaara A, Hayden J, Koes B (2007) Statistical significance versus clinical importance. *Spine* 32:1785–1790. doi:[10.1097/BRS.0b013e3180b9ef49](https://doi.org/10.1097/BRS.0b013e3180b9ef49)