The Spine Journal 19 (2019) 71–78

Clinical Study

Minimum clinically important difference of the health-related quality of life scales in adult spinal deformity calculated by latent class analysis: is it appropriate to use the same values for surgical and nonsurgical patients?

Selcen Yuksel, PhD[a], Selim Ayhan, MD[b,c], Vugar Nabiyev, MD[b], Montse Domingo-Sabat, PhD[d], Alba Vila-Casademunt, PhD[d], Ibrahim Obeid, MD[e], Francisco Sanchez Perez-Grueso, MD[f], Emre Acaroglu, MD[g,*], European Spine Study Group (ESSG) [d]

[a] *Department of Biostatistics, Yildirim Beyazit University, Ankara, Turkey*
[b] *ARTES Spine Center, Ankara, Turkey*
[c] *Vocational School of Health Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey*
[d] *Hospital Universitari Vall d'Hebron, Barcelona, Spain*
[e] *Bordeaux University Hospital, Bordeaux, France*
[f] *Hospital Universitari La Paz, Madrid, Spain*
[g] *Ankara Spine Center, Ankara, Turkey*
Received 30 January 2018; revised 4 July 2018; accepted 5 July 2018

Abstract

BACKGROUND CONTEXT: Health-related quality of life (HRQOL) parameters have been shown to be reliable and valid in patients with adult spinal deformity (ASD). Minimum clinically important difference (MCID) has become increasingly important to clinicians in evaluating patients with a threshold of improvement that is clinically relevant.

PURPOSE: To calculate MCID and minimum detectable change (MDC) values of total scores of the Core Outcome Measures Index (COMI), Oswestry Disability Index (ODI), Physical Component Summary (PCS), Mental Component Summary (MCS) of the Short Form 36 (SF-36), and Scoliosis Research Society 22R (SRS-22R) in surgically and nonsurgically treated ASD patients who have completed an anchor question at pretreatment and 1-year follow-up.

STUDY DESIGN/SETTING: Prospective cohort.

PATIENT SAMPLE: Surgical and nonsurgical patients from a multicenter ASD database.

OUTCOME MEASURES: Self-reported HRQOL measures (COMI, ODI, SF-36, SRS-22R, and anchor question).

METHODS: A total of 185 surgical and 86 nonsurgical patients from a multicenter ASD database who completed pretreatment and 1-year follow-up HRQOL scales and the anchor question at the first year follow-up were included. The anchor question was used to determine MCID for each HRQOL measure. MCIDs were calculated by an anchor-based method using latent class analysis (LCA) and MDCs by a distribution-based method.

RESULTS: All differences between means of baseline and first year postoperative total score measures for all scales demonstrated statistically significant improvements in the overall population as well as the surgically treated patients but not in the nonsurgical group. The calculated MDC and MCID values of HRQOL parameters in the entire study population were 1.34 and 2.62 for COMI, 10.65 and 14.31 for ODI, 6.09 and 7.33 for SF-36 PCS, 6.14 and 4.37 for SF-36 MCS, and 0.42 and 0.71 for SRS-22R. The calculated MCID values for surgical and non-surgical treatment groups

Author disclosures: ***SY:*** Nothing to disclose. ***SA:*** Nothing to disclose. ***VN:*** Nothing to disclose. ***MDS:*** Nothing to disclose. ***AVC:*** Grants; Depuy Synthes, (D) ***IO:*** Grants; Depuy Synthes, (D) ***FSPG:*** Grants; Depuy Synthes, (D) ***EA:*** Grants; Depuy Synthes, (C) ***ESSG:*** Grants; Depuy Synthes, (F) Medtronic, (D).

* Corresponding author. Ankara Spine Center, Iran Caddesi 45/2 Kavaklidere, Ankara 06700, Turkey. Tel.: (90) 312-4670442; fax: (90) 312-4673915.

E-mail address: acaroglue@gmail.com (E. Acaroglu).

https://doi.org/10.1016/j.spinee.2018.07.005

were 2.76 versus 1.20 for COMI, 14.96 versus 2.45 for ODI, 7.83 versus 2.15 for SF-36 PCS, 5.14 versus 2.03 for SF-36 MCS, and 0.94 versus 0.11 for SRS-22R; the MDC values for surgical and nonsurgical treatment groups were 1.22 versus 1.51 for COMI, 10.27 versus 9.45 for ODI, 5.16 versus 6.77 for SF-36 PCS, 6.05 versus 5.67 for SF-36 MCS, and 0.38 versus 0.43 for SRS-22R.
CONCLUSIONS: This study has demonstrated that MCID calculations for the HRQOL scales in ASD using LCA yield values comparable to other studies that had used different methodologies. The most important finding was the significantly different MCIDs for COMI, ODI, SF-36 PCS and SRS-22 in the surgically and nonsurgically treated cohorts. This finding suggests that a universal MCID value, inherent to a specific HRQOL for an entire cohort of ASD may not exist. Use of different MCIDs for surgical and nonsurgical patients may be warranted.

Keywords: Adult spinal deformity; COMI; MCID; MDC; ODI; Outcome; SF-36; SRS-22R; Surgery.

Introduction

Health-related quality of life (HRQOL) measures are used to evaluate the efficacy of treatment modalities in any disease including the adult spinal deformity (ASD). Previous studies have shown that Core Outcome Measures Index (COMI), Oswestry Disability Index (ODI), Short Form 36 (SF-36), and Scoliosis Research Society (SRS)-22 are reliable and valid in patients with ASD to measure HRQOL [1–4]. On the other hand, changes in HRQOL by time and/or treatment may not correlate with the direction (positive vs. negative) as well as the magnitude of clinical improvements in outcomes as they are perceived by the patients. So, it may be important to know whether a change in the score of an HRQOL assessment tool is perceived by the patients in the same direction as well as at a magnitude relevant to them.

A statistical way of demonstrating whether the perceived change is a clinically meaningful difference is the determination of the minimum clinically important difference (MCID). MCID was first defined by Jaeschke et al in 1989 as "the smallest difference in score in the domain of interest which patients perceive as beneficial" [5]. While definitions have varied greatly since then [6], the fundamental idea has remained the same: MCID is a calculated threshold value in an outcome of interest that patients and clinicians perceive as clinically meaningful, ie, a value that demonstrates an appreciable *change in outcome* [7]. There exists no ''gold standard'' methodology for estimating the MCID in the literature. Amongst numerous statistical methods, anchor-based and distribution-based calculations have been the most frequently used for conditions including ASD. There are several studies which had reported on MCID values for specific HRQOL measures using different calculation methodologies [8–13].

The purposes of this study are to calculate MCID values of COMI, ODI, SRS-22R, SF36-MCS, and SF36-PCS using an anchor-based analysis, to interpret the anchor-based MCID values with the distribution based MCID (MDC) values and iterate the potential differences and to analyze whether the method of treatment (surgical vs. nonsurgical) has any impact on these calculated values in ASD.

Materials and methods

Patients

Patients included in a multicenter prospective database of ASD were enrolled into this prospective study by filling out an anchor question that aimed to measure their perception of change in their condition by the administered treatment (or, thereby, in the absence of any treatment) at the first year [5]. This anchor question as well as most of the HRQOL tests was filled out by the patients themselves under the supervision of a clinical research assistant at each participating center. For the sake of simplicity, only patients with the most frequent diagnoses of degenerative and idiopathic ASD (which constitutes over 80% of the entire cohort) were included and those with other diagnoses of congenital, neuromuscular, post-traumatic were excluded. Patients who did not complete the anchor question at the 1-year follow-up were also excluded. In summary, the inclusion criteria were as follows:

1. Patients with ASD defined as follows:
   a. Any frontal plane deformity over 20°.
   b. Thoracic kyphosis over 60°.
   c. Spinal Vertical Axis (SVA) over 5 cm.
   d. Pelvic Tilt (PT) over 25°.
2. Consenting to be enrolled in the prospective database and being evaluated by the anchor question,
3. Primary diagnosis of degenerative or idiopathic deformity.
4. Complete 1-year follow-up anchor question data.

Of note, the decision for the type of treatment was based on the individual discretion of the treating physician in agreement with the patient and was not based on standardized criteria.

HRQOL measures and anchor question

The validated German, Spanish, French and Turkish [3] versions of the following four questionnaires were completed by the patients at baseline and at 1-year follow-up. Patients completed the *anchor question* just 1 year after.

1. *Core Outcome Measures Index (COMI)*: A short, validated, multidimensional outcome instrument to use measure health quality of patients with back problem. This seven-item, short and easy to use instrument has five domains: pain, back-related function, symptom-specific well-being, generic quality of life and work, and social disability in the previous month. Pain scores are indicated on a 0–10 numeric rating scale. Response categories for other items are five-point scales. Lower scores indicating better outcomes.
2. *Oswestry Disability Index (ODI)*: ODI has 10 items to measure back pain and nine different activities of daily life: personal care, lifting, walking, sitting, standing, sleeping, sex life, social life and travelling. Each item is scored from 0 to 5, with higher values representing greater disability. The total score is multiplied by 2, to give a score from 0 to 100.
3. *Short Form 36 (SF-36)*: This health survey contains 36 items scattered into eight domains: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health. For each domain, a score ranging from 0 (worst measured health) to 100 (best measured health) is calculated.
4. *Scoliosis Research Society-22R (SRS-22R)*: A scoliosis specific scale contains 22 items and 5 domains: function/activity, pain, self-image, mental health, satisfaction. Each domain score ranges from 1 to 5, with higher total scores indicating better outcomes.
5. *Anchor question*: This scale measures a patient's improvement or deterioration over time after treatment just by using one single question translated into the different languages corresponding to the language in the countries they live in. The back translations of this question (to English from the original language it was asked in) can be listed as:

- Compared to one year ago, how are you in regard to your back problems?
- Compared to one year ago, how are your back problems?
- Compared to your situation before the operation, how has the spine surgery modified your back condition?
- Compared to one year ago, how are you from your back problems?

Response items range between (−) 7 (a very great deal worse) and (+) 7 (a very great deal better), on a 15-point Likert type scale. This anchor question was used as an anchor to determine the MCID in this study.

Anchor-based MCID calculation

Anchor question answers were used as anchors to evaluate each HRQOL scale. Although there are various established threshold definitions to split anchor question answers as "unsatisfied" and "satisfied" in literature, in this study, a latent class analysis (LCA) was used to discriminate between those perceiving an improvement as opposed to those who do not (i.e. worse or unchanged) in order to avoid any subjectivity and to obtain results with robust theoretical basis. For this aim, response patterns in scales were taken as covariates in addition to anchor question responses that were taken as indicators in the LCA model. This model was fixed at a 2-class solution. To evaluate model fit within this two class solution, classification error and pseudo R^2 statistics were used (Table 1). Values of smaller classification errors (close to 0.0) and of pseudo R^2 statistics as close to 1.0 as possible were accepted as ideal. Thresholds calculated from the LCA model as described can be found in Table 1. R^2 statistics ranging between 0.78 and 0.99 demonstrates that 2-class solutions with given thresholds has

Table 1
Latent class analysis 2-class solution results for each HRQOL scale

Overall	COMI	ODI	SF-36 PCS	SF-36 MCS	SRS 22R
Classification error	0.01	0.06	0.01	0.04	0.002
Pseudo R^2 statistics	0.91	0.79	0.94	0.79	0.99
Threshold*	1	4	3	0	1
Surgical group					
Classification error	0.07	0.04	0.01	0.02	0.002
Pseudo R^2 statistics	0.78	0.87	0.96	0.93	0.99
Threshold*	3	4	4	4	5
Nonsurgical group					
Classification error	0.009	0.001	0.003	0.01	0.0013
Pseudo R^2 statistics	0.95	0.99	0.98	0.95	0.99
Threshold*	−2	−3	−2	−2	−1

* Threshold values calculated by using LCA to split patients as satisfied and unsatisfied.

Table 2
Linear regression model fits to evaluate the effects; the effect of treatment modality on the average change score in patients' ratings. The first row indicates dummy coding consistent with the presence of a categorical independent variable in the linear regression model

Independent variable	Outcome variable	ANOVA model significance
	Δ_{COMI}	0.095
	Δ_{ODI}	0.075
	$\Delta_{SF36\ PCS}$	0.020
	$\Delta_{SF36\ MCS}$	0.209
Treatment (surgical (1), nonsurgical (0))	$\Delta_{SRS\ 22}$	0.001

Δ = Baseline total score − 1-year postoperative total score.

indeed split the data in two distinct groups of perceived improvement or not satisfactorily.

As MCID reflects the magnitude of change in the measure associated with an arbitrary definition of the smallest important change in an anchor question, it may be estimated by the average change score in patients rating some improvement on scales. Based on this, the anchor-based MCID estimate was calculated as the mean change score on scales, corresponding to patients who have anchor question responses higher than the established threshold value calculated by the LCA. In addition, by fitting a linear regression model, the effect of treatment modality on the average change score in patients' ratings were analyzed. These calculations can be seen on Table 2.

Distribution-based calculation of MDC

Distribution based method is based on the statistical distributions of change scores and proportions of the pooled standard deviation (SD_{pooled}). The MDC is the minimal amount of change score outside of measurement error that may reflect a true change. It is calculated by multiplying the standard error of measurement (SEM) by the *z score* associated with the desired confidence level and the square root of 2, adjusting for sampling from 2 different measures. SEM quantifies within-subject variability and reflects the amount of measurement error. SEM is estimated as the SD_{pooled} of pre- and post-treatment assessments multiplied by the square root of $(1-r)$, where r is the intraclass correlation coefficient (ICC) [14]. Under the assumption of "patients and error terms are random and independent", a two-way random model was used to calculate ICC. For the ICC ranges in this study, MDCs were calculated based on a 90% confidence interval.

To compare baseline and first year postoperative total score measures, paired t-test or Wilcoxon signed rank tests were used in accordance with the distribution of the difference of two dependent measures. To compare unsatisfied and satisfied groups based on the first postoperative year total scores, t-test was used if scores were normally distributed, Mann-Whitney U test was used otherwise to compare two independent groups. All statistical analyses were done using R i386.3.2.1 language [15]. *"Lca"* package in R was used for LCA analysis. Type I error rate was taken as α=0.05 for statistical significance.

Results

A total of 185 surgical and 86 nonsurgical patients from the database were included for MCID calculations. The study was performed in 80 (43.2%) degenerative and 105 idiopathic (56.8%) surgically treated, and 9 degenerative (10.5%) and 60idiopathic (69.8%) nonsurgical patients with ASD; whereas there were 17 patients (19.8%) with a missing value in the diagnosis domain (which means that the specific diagnosis of this patient has not been entered into the database by the center) in the nonsurgical group, who were nevertheless included in order to increase the number of patients enrolled in the nonsurgical group. Mean age was 52.4±20 and 44.9±16.9 years for surgical and nonsurgical patients, respectively. A total of 188 patients had completed COMI, 257 patients had completed ODI, 247 patients had completedSF-36 MCS and SF-36 PCS, and 255 patients had completed SRS-22R. Median values of responses received from the anchor question for surgical patients is 5 (min=(−)7; max=(+)7; IQR=3) and nonsurgical patients is 0 (min=(−)7; max=(+)7; IQR=2). Overall median of anchor question was 3 (min=(−)7; max=(+)7; IQR=6).

All differences between means of baseline and first year post-operative total score measures for all scales demonstrated statistically significant improvements in the overall population as well as the surgically treated patients, but not in the nonsurgical group (Table 3). The differences in the mean total scores of each scale between the "perception of improvement" and "not" derived from using threshold values were also statistically significant (Table 4). MCID values for overall and subgroups of each scale are summarized in Table 5. As seen in Table 5, SEM, MDC and anchor-based MCID values are different for the overall population and the treatment groups.

Discussion

This study aimed to calculate and report the MCID and MDC values of commonly used HRQOL scales (COMI, ODI, SF-36 PCS, SF-36 MCS, SRS-22R) in ASD. In addition, these values were calculated in two separate cohorts of surgically and nonsurgically treated patients to see whether MCID and MDC would be sensitive to the type of treatment. Our results suggest that MCIDs calculated using anchor-based LCA methodology are fairly in line with those reported in the literature before. The MDC values on the other hand, were calculated to be substantially different than the MCIDs for some parameters. Further, our results also demonstrate that the MCID and MDC values for a given HRQOL parameter can be sensitive to the type of treatment undertaken, ie, surgical versus nonsurgical.

Table 3
Comparison of baseline and 1-year-postoperative scores for each HRQOL scale

Scale	N*		Baseline Mean Total Score (SD)	1-yr postoperative Mean Total Score (SD)	p
COMI	188	Overall	5.4 (2.7)	3.9 (2.6)	<.001
	109	Surgical	6.3 (2.2)	3.8 (2.4)	<.001
	79	Nonsurgical	4.1 (2.7)	4.0 (2.8)	.439
ODI	257	Overall	35.8 (21.3)	27.9 (18.6)	<.001
	172	Surgical	40.8 (20.8)	28.0 (17.8)	<.001
	85	Nonsurgical	25.7 (2.0)	27.4 (20.2)	.182
SF-36 PCS	247	Overall	37.9 (10.4)	41.8 (10.2)	<.001
	162	Surgical	36.0 (8.7)	41.9 (8.9)	<.001
	85	Nonsurgical	41.4 (12.2)	41.4 (12.3)	.995
SF-36 MCS	247	Overall	43.6 (12.4)	46.8 (11.0)	<.001
	162	Surgical	41.5 (11.9)	45.9 (11.3)	<.001
	85	Nonsurgical	47.5 (12.3)	48.6 (10.5)	.205
SRS-22R	255	Overall	3.0 (0.8)	3.5 (0.7)	<.001
	170	Surgical	2.9 (0.7)	3.5 (0.7)	<.001
	85	Nonsurgical	3.5 (0.8)	3.4 (0.8)	.224

* The reason for surgical and nonsurgical total patient numbers not being equal to overall patient number within satisfied and unsatisfied groups is the differences in the threshold values from LCA for each scale (see Table 1),which was used as the base of inclusion in this table.

Minimum clinically important difference

Establishment of well-defined and validated MCID values for any clinical measure should be of utmost importance. Without any information on MCID values, evaluation of treatment results need necessarily be based only on the statistical significance of the changes mediated by the treatment, which may not be able to provide any information on the clinical relevance of the changes accrued. Hence, reporting treatment results based on their relationship with the corresponding MCID values are on their way to become the gold standard in clinical research and reporting. This study is not the only study that had aimed and defined MCID values for several HRQOL measures in ASD. There are two main differences between this and other studies though. First, this is the only study that had analyzed the MCID for the five most frequently used scales using the same database compared to others reporting results on one specific or a couple of scales only [9–11]. We believe this multidimensionality in addition to the overall similarity of our results to what has been reported before adds to the validity of our results.

Second, this is the only study that has calculated the MCID values using anchor-based calculation methodology and LCA for a population of ASD, using an external anchor question. To date, there are three separate studies reporting on MCIDs using three different methodologies. Copay and coworkers have used the distribution based methodology described above for the calculation of MDC and suggested that (for this specific condition) MDC should be essentially synonymous to MCID [9]. In the study by Gum and coworkers, MCID values were identified using receiver

Table 4
Comparison of unsatisfied and satisfied groups by mean change on total score of each HROOL scale.

Scale		Satisfied Mean Change (SD-n)	Unsatisfied Mean Change (SD-n)	p
COMI	Overall	2.6 (2.7–97)	0.3 (1.8–91)	<.001
	Surgical	2.8 (2.8–76)	1.6 (2.3–33)	.036
	Nonsurgical	0.4 (1.6–58)	−0.6 (1.4–21)	.007
ODI	Overall	14.3 (17.6–99)	4.0 (14.2–158)	<.001
	Surgical	15.0 (17.6–95)	10.0 (13.8–77)	.045
	Nonsurgical	−0.4 (10.4–67)	−6.5 (15.9–18)	.053
SF-36 PCS	Overall	7.3 (8.3–119)	0.6 (7.3–128)	<.001
	Surgical	7.8 (8.7–88)	3.6 (7.3–74)	.001
	Nonsurgical	1.2 (6.7–61)	−3.0 (8.0–24)	.020
SF-36 MCS	Overall	4.4 (10.3–186)	0.2 (8.4–61)	.002
	Surgical	5.1 (11.2–88)	3.4 (10.4–74)	.323
	Nonsurgical	2.0 (8.0–61)	−1.3 (6.6–24)	.071
SRS-22R	Overall	0.7 (0.6–148)	0.03 (0.5–107)	<.001
	Surgical	0.9 (0.6–66)	0.5 (0.6–104)	<.001
	Nonsurgical	−0.03 (0.4–50)	0.2 (0.4–35)	.017

Table 5
Distribution based and anchor-based MCID values of each HRQOL scale for subgroups.

Scale		SEM	MDC	Anchor-based MCID
COMI	Overall	0.48	1.34	2.62
	Surgical	0.44	1.22	2.76
	Nonsurgical	0.55	1.51	1.20
ODI	Overall	3.84	10.65	14.31
	Surgical	3.71	10.27	14.96
	Nonsurgical	3.41	9.45	2.45
SF-36 PCS	Overall	2.20	6.09	7.33
	Surgical	1.86	5.16	7.83
	Nonsurgical	2.44	6.77	2.15
SF-36 MCS	Overall	2.21	6.14	4.37
	Surgical	2.18	6.05	5.14
	Nonsurgical	2.05	5.67	2.03
SRS 22R	Overall	0.15	0.42	0.71
	Surgical	0.14	0.38	0.94
	Nonsurgical	0.16	0.43	0.11

operating characteristic (ROC) curves [11]. In this methodology, analysis comprises of identifying a meaningful cut-off *"value in a ROC curve corresponding to the point of equal trade-off between greatest sensitivity and specificity to distinguish the 'worse' from the 'about the same'."* This is essentially different from the other methodologies used in that as it aims to identify the difference between "worse" and "about the same," it does not necessarily measure the smallest measurable increment. Finally, Crawford and coworkers have used similar methodology to what has been used in this study; the differences being their selection of the anchor from within the SRS-22R score questions rather than using an external anchor [10] and that they did not calculate thresholds in the responses based on response patterns but rather assigned the neutral answer as a threshold. Their calculated MCID value for the overall SRS-22R score for a surgically treated group of patients is exactly the same (0.71) with our overall population value and significantly different than that of our surgically treated patient group (0.94). This difference is apparently caused by the different methodologies being used. In this regard, we tend to believe that our result may be more accurate as an external anchor (vs. an internal anchor which is a component of the scale that is analyzed, i.e., the satisfaction question of SRS-22R) may not only validate but also increase the legitimacy of MCID calculations considerably.

Minimum detectable change

MDC by itself, unlike MCID, may not necessarily be an important clinical parameter that may be instrumental in comparing treatments and outcomes. Still, we wanted to calculate the MDC values for the scales we have been calculating the MCID, for two reasons:

1. Calculation of MDC may serve as an internal control of the data set used for the calculation of MCID. In cases in which the MDC and MCID values are considerably different, we may start questioning the skewness and kurtosis of distribution of the dataset for such an analysis. Although the results of this study have demonstrated some differences, these differences have not been more than one standard deviation for any scale or group (Table 5).
2. From a clinical point of view, in instances which the MCID is smaller in magnitude compared to the MDC. The MDC, by definition, should be used in MCID's stead. The opposite would virtually imply that, for at least the data set we have used in this study, we would be capable of calculating the MCID but incapable of detecting it. In our result set, this condition is relevant for COMI (nonsurgical), ODI (nonsurgical), SF-36 PCS (nonsurgical), SF-36 MCS (overall, surgical and nonsurgical) and SRS-22R (nonsurgical). This discrepancy may potentially be secondary to the relatively lower values of MCID in the nonsurgical group of patients.

MCID's sensitivity to treatment methodology

As also eluded to above, the MCID calculations in the nonsurgical group is considerable lower in values compared to those patients who had been treated surgically. We must iterate at this point that it is virtually impossible to know whether these differences are statistically significant as there are no defined methodologies for comparing calculated MCIDs with each other. On the other hand, they are fairly large and need to be accepted as being clinically relevant (as an example, the MCID of ODI in the nonsurgical group was calculated to be 2.45). There may be two reasons for this sensitivity of the MCID (exclusively) to the treatment. One that readily springs to mind is that this finding may represent an artifact, potentially due to the relatively smaller sample size and/or an abnormal distribution of the answers to the anchor question. Upon further analysis of these factors, we concluded that both factors might have influenced our results. It is true that the sample size of the nonsurgical group is smaller than that of the surgical, and it is also quite remarkable that the answers to anchor question in this group were found to be more normally distributed as opposed to answers tending to be clustered at the extremes in the surgical group (more so at the positive end, as demonstrated by the median value of 5 in this group). This latter tendency was not substantiated for the purposes of this study but given the relative ineffectiveness of nonsurgical treatment in ASD that has been demonstrated by several studies [16,17]. A "normal" distribution seems logical.

As a second and more plausible alternative though, this finding may as well be legitimate and free of any major artifact. This line of thought is based on a study by Carreon and coworkers demonstrating that the MCID values in primary and revision back surgery cohorts may be different [18] (albeit not substantially). Further, Parker and coworkers have demonstrated that MCID calculations yielded

substantially different MCID values for four different methods (average change; minimum detectable change; change difference; and receiver operating characteristic curve (ROC) analysis for 2 separate anchors of health transition item of the SF-36 and satisfaction index) used in two different studies on two separate populations [19,20]. Although these two populations (patients with surgically treated proximal junctional kyphosis in one and surgically treated pseudoarthrosis in the other) are different than our population and not necessarily represent a pure ASD population, these studies collectively suggest that MCID calculations may be prone to significant variations based not only on the calculation methodology used but also on other parameters including diagnosis.

This study has several shortcomings. One is the not-so-standard translation of the anchor question into different languages (and back) that might have affected the response patterns to that question radically. Although we do recognize this as a potential shortcoming, we do not think that it had radically affected the results of this study for two reasons. Firstly, questions posed in different countries are not exact translations of a standard phrase but do have similar meanings. This is probably an inherent characteristic of a multinational study, translation may need to be adapted to the structure of the given language. Secondly, this variation and uncertainty of response patterns to anchor question is the very reason why LCA was used in this study. As has been iterated to above, different thresholds based on variations in response patterns (not necessarily because of but including cultural and/or language differences) have been identified and used to minimize any effect of potentially different perceptions of anchor question by different populations.

A potential second shortcoming may be the inclusion of 17 patients for whom the diagnosis had not been entered into the database. This may have been a major shortcoming had patients been classified based on their diagnoses but since any calculation or discussion in regards to this parameter was performed or used, it may as well be considered as a very minor shortcoming.

Finally, a third shortcoming may be the utilization of LCA in MCID calculations for the first time in literature. Although we think that this has provided advantages that had been discussed above, this methodology may need to be further tested in other studies as well as fields before being indisputably justified.

Conclusions

This study has described the MCID and MDC values for surgically and nonsurgically treated ASD patients, pooled and separately. Of significance, it was demonstrated that the anchor-based MCID values were considerably lower in values for the nonsurgical patients compared to surgically treated patients. This finding strongly suggests that a single constant MCID value which is applicable to an entire cohort of adult spinal deformity patients may not exist.

Acknowledgments

This study had been done using resources form the European Spine Study Group. European Spine Study Group is funded by a research grant from Depuy Synthes Spine.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.spinee.2018.07.005.

References

[1] Bridwell KH, Cats-Baril W, Harrast J, Berven S, Glassman S, Farcy JP, et al. 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 (Phila Pa 1976) 2005 Feb 15;30(4):455–61. PubMed PMID:15706344.

[2] Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy 1980;66(8):271–3.

[3] Mannion AF, Vila-Casademunt A, Domingo-Sàbat M, Wunderlin S, Pellisé F, Bago J, et al. The Core Outcome Measures Index (COMI) is a responsive instrument for assessing the outcome of treatment for adult spinal deformity. Eur Spine J 2016 Aug;25(8):2638–48 PubMed PMID: 26519374. doi: 10.1007/s00586-015-4292-4.

[4] Walsh TL, Hanscom B, Lurie JD, Weinstein JN. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. Spine 2003;28(6):607–15.

[5] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controll Clin Trials 1989;10(4):407–15.

[6] Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Curr Opin Rheumatol 2002;14(2):109–14.

[7] Chung AS, Copay AG, Olmscheid N, Campbell D, Walker JB, Chutkan N. Minimum clinically important difference: current trends in the spine literature. Spine 2017;42(14):1096–105.

[8] International Spine Study Group, McCarthy I, Bess S, Line B, O'Brien M, Schwab F J, et al. Calculating and Defining Minimally Important Clinical Difference (MCID) and Substantial Clinical Benefit (SCB) Values for Adult Spinal Deformity (ASD): A Robust Methodology for Consistent Data Reporting. Spine J 2013;13(9). 75S.

[9] Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. Spine J 2008;8(6):968–74.

[10] Crawford 3rd CH, Glassman SD, Bridwell KH, Berven SH, Carreon LY. The minimum clinically important difference in SRS-22R total score, appearance, activity and pain domains after surgical treatment of adult spinal deformity. Spine 2015;40(6):377–81.

[11] Gum JL, Glassman SD, Carreon LY. Clinically important deterioration in patients undergoing lumbar spine surgery: a choice of evaluation methods using the Oswestry Disability Index, 36-Item Short Form Health Survey, and pain scales: clinical article. J Neurosurg Spine 2013;19(5):564–8.

[12] Bago J, Perez-Grueso FJ, Les E, Hernandez P, Pellise F. Minimal important differences of the SRS-22 Patient Questionnaire following

surgical treatment of idiopathic scoliosis. Eur Spine J 2009;18 (12):1898–904.

[13] Wells G, Beaton D, Shea B, Boers M, Simon L, Strand V, et al. Minimal clinically important differences: review of methods. The Journal of Rheumatology February 2001;28(2):406–12.

[14] de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59 (10):1033–9.

[15] Team RDC. R: A language and environment for statistical computing. http://www.R-project.org; 2008. Accessed August 1, 2017.

[16] Acaroglu E, Yavuz AC, Guler UO, Yuksel S, Yavuz Y, Domingo-Sabat M, et al. A decision analysis to identify the ideal treatment for adult spinal deformity: is surgery better than non-surgical treatment in improving health-related quality of life and decreasing the disease burden? Eur Spine J 2016 Aug;25(8):2390–400 PubMed PMID: 26821143. doi: 10.1007/s00586-016-4413-8.

[17] Bridwell KH, Glassman S, Horton W, Shaffrey C, Schwab F, Zebala LP, et al. Does treatment (nonoperative and operative) improve the two-year quality of life in patients with adult symptomatic lumbar scoliosis: a prospective multicenter evidence-based medicine study. Spine (Phila Pa 1976) 2009 Sep 15;34(20):2171–8 PubMed PMID: 19752703. doi: 10.1097/BRS.0b013e3181a8fdc8.

[18] Carreon LY, Bratcher KR, Canan CE, Burke LO, Djurasovic M, Glassman SD. Differentiating minimum clinically important difference for primary and revision lumbar fusion surgeries. J Neurosurg Spine 2013;18(1):102–6.

[19] Parker SL, Mendenhall SK, Shau D, Adogwa O, Cheng JS, Anderson WN, et al. Determination of minimum clinically important difference in pain, disability, and quality of life after extension of fusion for adjacent- segment disease. J Neurosurg Spine 2012 Jan;16(1):61–7 PubMed PMID: 21962034. doi: 10.3171/2011.8.SPINE1194.

[20] Parker SL, Adogwa O, Mendenhall SK, Shau DN, Anderson WN, Cheng JS, et al. Determination of minimum clinically important difference (MCID) in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis. Spine J 2012 Dec;12 (12):1122–8 PubMed PMID: 23158968. doi: 10.1016/j.spinee.2012.10.006.