The SPINE JOURNAL

The Spine Journal 000 (2024) 1−14

Clinical Study

Machine learning clustering of adult spinal deformity patients identifies four prognostic phenotypes: a multicenter prospective cohort analysis with single surgeon external validation

Sarthak Mohanty, BS[a], Fthimnir M. Hassan, MPH[a,*], Lawrence G. Lenke, MD[a], Erik Lewerenz, BS[a], Peter G. Passias, MD[b], Eric O. Klineberg, MD[c], Virginie Lafage, PhD[d], Justin S. Smith, MD, PhD[e], D. Kojo Hamilton, MD[f], Jeffrey L. Gum, MD[g], Renaud Lafage, MS[d], Jeffrey Mullin, MD[h], Bassel Diebo, MD[i], Thomas J. Buell, MD[f], Han Jo Kim, MD[j], Khalid Kebaish, MD[k], Robert Eastlack, MD[l], Alan H. Daniels, MD[i], Gregory Mundis, MD[l], Richard Hostin, MD[m], Themistocles S. Protopsaltis, MD[b], Robert A. Hart, MD[n],

FDA device/drug status: Not applicable.

Author disclosures: ***SM:*** Nothing to disclose. ***FMH:*** Nothing to disclose. ***LGL:*** Grant: AO Spine, International Spine Summit Group, Setting Scoliosis Straight Foundation (monies paid to institution). ***EL:*** Nothing to disclose. ***PGP:*** Consulting: Medtronic (B); Globus (B); Cerapedics (B). ***EOK:*** Royalites: Stryker Spine (B). Stock Ownership: MMI (A); relatable (B). Endowments: UTHealth Houston (Paid to Institution). Fellowship Support: SAONA Spine, Medtronic (monies paid to institution) ***VL:*** Royalties: Nuvasive (C). Stock Ownership: VFT Solutions LLC (50%); SeeSpine LLC (33%). Consulting: Alphatec (D); Globus Medical (D). Speaking and/or Teaching Arrangements: J&J (C); Stryker (C). Scientific Advisory Board/Other Office: ISSG (n/a). ***JSS:*** Grant: ISSGF/ DePuys Synthes (B). Royalties: Zimvie (F); NuVasive (B); Thieime (A). Consulting: Zimvie (B); NuVasive (B); SeaSpine (B); Cerapedics (B); Carlsmed (C). Research Support (Investigator Salary, Staff/Materials: ISSGF/DePuy Synthes (paid to institution). Fellowship Support: AOSpine (paid to institution). ***DKH:*** Nothing to disclose. ***JLG:*** Grant: Alan L. & Jacqueline B. Stuart Spine Center, Biom'up, Cerapedics, Inc, Empirical Spine, Inc, Medtronic, National Spine Health Foundation, Scoliosis Research Society, Stryker, International Spine Study Group (monies paid to institution). Consulting fee or honorarium: Depuy (B), Acuity (D); Medtronic (F); Nuvasive (D); Stryker (C); FYR Medical (B); Expanding Innovations (B); NASS (B): Broadwater (B). Royalties: Acuity (E); Medtronic (F); Nuvasive (F). Stock Ownership: Cingulate Inc. (E). Speaking and/or Teaching Arrangements: Kyana (A). Trips/Travel: Fischer Owens Fund (B). Scientific Advisory Board/Other Office: Medtronic (F); National Spine Health Foundation (B); FYR Medical (B). ***RL:*** Consulting: CarlsMed (B). ***JM:*** Nothing to disclose. ***BD:*** Nothing to disclose. ***TJB:*** Nothing to disclose. ***HJK:*** Grant: ISSGF (D). Consulting Fee or honorarium: (B). Royalties: immerbiomet, K2M/Stryker, Acuity Surgical (monies paid to institution). Fellowship: Nuvasive (paid to institution). ***KK:*** Royalties: Depuy Synthes (F); Stryker (F); Orthofix (D); SpineCraft (C). Consulting: Depuy Synthes (E). ***RE:*** Royalties: Nuvasive (F); SeaSpine (E); SI Bone (E); Aesulap (C); Globus (B). Consulting: Nuvasive (B); SI Bone (C); SeaSpine (B); Medtronic (B); Neo (B); Silony (B); Spinal Elements (C). ***AHD:*** Royalties: Spineart (F); Stryker (D). Consulting: Medtronic (D). Research Support (Investigator Salary, Staff/Materials): Orthofix (D). Fellowship Support: Medtronic (E). ***GM:*** Consulting fee of honorarium: Carlsmed: (B); Nuvasive (F); SeaSpine (D); SI Bone (B); Viseon (B). Royalties: Nuvasive (E); SeaSpine (E); Stryker (E). Consulting: Carlsmed (B); Nuvasive (F); SeaSpine (D); SI Bone (B); Viseon (B). Grants: Nuvasive (D); Medtronic (D); SeaSpine (E). Fellowship Support: AO (E); Nuvasive (E); SeaSpine (E). ***RH:*** Nothing to disclose. ***TSP:*** Royalties: Altus (C, Paid directly to institution/employer); Private Investments: OnPoint Surgical (F, Paid directly to institution/employer); Consulting: Globus, Nuvasive Medtronic, Stryker (D), Paid directly to institution/employer); Grants: Globus (E, Paid directly to institution/employer); Fellowship Support: Globus (E), Nuvasive Medtronic (F, Paid directly to institution/employer). ***RAH:*** Royalties: Globus (A); SeaSpine (B). Stock Ownership: MiRus; Amplify. Consulting: Globus: (A); Medtronic: (B); SeaSpine (B); MiRus (B). Trips/Travel: SeaSpine (B); Atex (B). ***MG:*** Grant: ISSGF (E). ***FJS:*** Royalties: Stryker (F); MSD (D); Zimmer Biomet (F). Consulting: Mainstay Medical (E); Stryker (D); Zimvie (D). ***CIS:*** Royalties: Nuvasive (D). Stock Ownership: Nuvasive (36,000 shares). Consulting: Proprio (B); Medtronic (C); Si-Bone (D) ***CPA:*** Royalties: Stryker (F); Biomet Zimmer Spine (C); Depuy Synthes (B); Nuvasive (B); Next Orthosurgical (F); Medicrea (B); K2M (None). Consulting: Depuy Synthes(B); Medtronic (B); Medicrea (B); K2M (C); Agada Medical (B); Carlsmed (B). Research Support (Investigator Salary, Staff/Materials): Titan Spine (E); DePuy Synthes (None); ISSG (C). ***DB:*** Royalties: DePuy Synthes (B): Globus (None); Blue Ocean Spine (None). Stock Ownership: (B). Consulting: Blue Ocean Spine (None); DePuy Synthes (A); Globus (A). Board of Directors: Scoliosis Research Society (None); International Spine Study Group (None). Research Support (Investigator Salary, Staff/Materials): DePuy Spine (C); Pfizer (B); Bioventus (B); International Spine Study Group (C). ***SB:*** Grant: Medtronic (F): Globus (E); Nuvasive (E); Stryker (E); Carlsmed (E). Royalties: Stryker (E); Nuvasive (B). Consulting: Atec (C). Speaking and/or Teaching Arrangements: Atec (B). Research Support (Investigator Salary, Staff/ Materials): DePuy Synthes (F); Nuvasive (E).

*Corresponding author. The Daniel and Jane Och Spine Hospital, New York Presbyterian, Columbia University Medical Center, 5141 Broadway, New York, NY, 10034, USA.

E-mail address: fh2444@cumc.columbia.edu (F.M. Hassan).

https://doi.org/10.1016/j.spinee.2024.02.010

1529-9430/

Munish Gupta, MD, MBA[o], Frank J. Schwab, MD[d], Christopher I. Shaffrey, MD[p], Christopher P. Ames, MD[q], Douglas Burton, MD[r], Shay Bess, MD[s], International Spine Study Group [a]

[a] Department of Orthopaedics, Columbia University Medical Center, New York, NY, USA
[b] Department of Orthopaedic Surgery, New York University Langone Medical Center, New York, NY, USA
[c] Department of Orthopaedic Surgery, University of California Davis Medical Center, Sacramento, CA, USA
[d] Department of Orthopaedic Surgery, Northwell Health Lenox Hill, New York, NY, USA
[e] Department of Neurosurgery, University of Virginia Medical Center, Charlottesville, VA, USA
[f] Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA, USA
[g] Department of Orthopaedic Surgery, Norton Leatherman Spine Center, Louisville, KY, USA
[h] Department of Neurosurgery, University at Buffalo, Buffalo, NY, USA
[i] Department of Orthopaedic Surgery, University Orthopedics, Providence, RI, USA
[j] Department of Orthopaedic Surgery, Hospital for Special Surgery, New York, NY, USA
[k] Department of Orthopaedic Surgery, John Hopkins Medical Institute, Baltimore, MD, USA
[l] Division of Orthopaedic Surgery, Scripps Clinic, La Jolla, CA, USA
[m] Department of Orthopaedic Surgery, Southwest Scoliosis and Spine Institute, Dallas, TX, USA
[n] Department of Orthopaedic Surgery, Swedish Neuroscience Institute, Seattle, WA, USA
[o] Department of Orthopaedic Surgery, Washington University School of Medicine, St. Louis, MO, USA
[p] Department of Neurosurgery, Duke University Medical Center, Durham, NC, USA
[q] Department of Neurosurgery, University of California San Francisco Spine Center, San Francisco, CA, USA
[r] Department of Orthopaedic Surgery, University of Kansas Medical Center, Kansas City, KS, USA
[s] Department of Orthopaedic Surgery, Denver International Spine Center, Denver, CO, USA

Received 9 August 2023; revised 11 January 2024; accepted 8 February 2024

Abstract

BACKGROUND CONTEXT: Among adult spinal deformity (ASD) patients, heterogeneity in patient pathology, surgical expectations, baseline impairments, and frailty complicates comparisons in clinical outcomes and research. This study aims to qualitatively segment ASD patients using machine learning-based clustering on a large, multicenter, prospectively gathered ASD cohort.

PURPOSE: To qualitatively segment adult spinal deformity patients using machine learning-based clustering on a large, multicenter, prospectively gathered cohort.

STUDY DESIGN/SETTING: Machine learning algorithm using patients from a prospective multicenter study and a validation cohort from a retrospective single center, single surgeon cohort with complete 2-year follow up.

PATIENT SAMPLE: About 805 ASD patients; 563 patients from a prospective multicenter study and 242 from a single center to be used as a validation cohort.

OUTCOME MEASURES: To validate and extend the Ames-ISSG/ESSG classification using machine learning-based clustering analysis on a large, complex, multicenter, prospectively gathered ASD cohort.

METHODS: We analyzed a training cohort of 563 ASD patients from a prospective multicenter study and a validation cohort of 242 ASD patients from a retrospective single center/surgeon cohort with complete two-year patient-reported outcomes (PROs) and clinical/radiographic follow-up. Using k-means clustering, a machine learning algorithm, we clustered patients based on baseline PROs, Edmonton frailty, age, surgical history, and overall health. Baseline differences in clusters identified using the training cohort were assessed using Chi-Squared and ANOVA with pairwise comparisons. To evaluate the classification system's ability to discern postoperative trajectories, a second machine learning algorithm assigned the single-center/surgeon patients to the same 4 clusters, and we compared the clusters' two-year PROs and clinical outcomes.

RESULTS: K-means clustering revealed four distinct phenotypes from the multicenter training cohort based on age, frailty, and mental health: Old/Frail/Content (OFC, 27.7%), Old/Frail/Distressed (OFD, 33.2%), Old/Resilient/Content (ORC, 27.2%), and Young/Resilient/Content (YRC, 11.9%). OFC and OFD clusters had the highest frailty scores (OFC: 3.76, OFD: 4.72) and a higher proportion of patients with prior thoracolumbar fusion (OFC: 47.4%, OFD: 49.2%). ORC and YRC clusters exhibited lower frailty scores and fewest patients with prior thoracolumbar procedures (ORC: 2.10, 36.6%; YRC: 0.84, 19.4%). OFC had 69.9% of patients with global sagittal deformity and the highest T1PA (29.0), while YRC had 70.2% exhibiting coronal deformity, the highest mean coronal Cobb Angle (54.0), and the lowest T1PA (11.9). OFD and ORC had similar alignment phenotypes with intermediate values for Coronal Cobb Angle (OFD: 33.7; ORC: 40.0) and T1PA (OFD: 24.9; ORC: 24.6) between OFC (worst sagittal alignment) and YRC (worst coronal alignment). In the single surgeon validation cohort, the OFC cluster experienced the greatest increase in SRS Function scores (1.34 points, 95%CI 1.01−1.67) compared to OFD (0.5 points, 95%CI 0.245−0.755), ORC (0.7 points, 95%CI 0.415−0.985), and YRC (0.24 points, 95%CI -0.024−0.504) clusters. OFD cluster patients improved the least over 2 years. Multivariable Cox regression analysis demonstrated that the OFD cohort had significantly worse reoperation outcomes compared to other clusters (HR: 3.303, 95%CI: 1.085−8.390).

CONCLUSION: Machine-learning clustering found four different ASD patient qualitative phenotypes, defined by their age, frailty, physical functioning, and mental health upon presentation, which primarily determines their ability to improve their PROs following surgery. This reaffirms that these qualitative measures must be assessed in addition to the radiographic variables when counseling ASD patients regarding their expected surgical outcomes.

Keywords: Adult spinal deformity; Classifications; Frailty; Machine learning; Mental health; Patient reported outcomes; Phenotypes; Spinal deformity surgery

Introduction

Adult spinal deformity (ASD) impacts approximately 60% of elderly patients, significantly affecting their health-related quality of life (HRQoL) [1−3]. Prospective evaluation of SF-36 scores indicate that patients with sagittal malalignment endure disability levels comparable to individuals with limited extremity use [1], surpassing other chronic conditions [4].

Although corrective surgery has proven beneficial for a significant proportion of ASD patients [5−9], heterogeneity in patient pathology [6], surgical expectations [10,11], baseline physical and mental impairments [11,12], and frailty [13] complicates precise clinical decision-making and patient counseling. For example, Smith et al. reported that while perioperative complication rates increased with age (17%, 42%, and 71% for ages 25−44, 45−64, 65−85 respectively), older patients experienced superior improvements in ODI and leg pain compared to younger individuals [6].

Over the last 2 decades, classification schemes like the Scoliosis Research Society (SRS)-Schwab classification have emerged to mitigate heterogeneity by conceptualizing uniform subgroups with distinct presentations, needs, and postoperative trajectories [14−17]. However, a classification scheme based solely on radiographic parameters correlated with PROMs may not adequately capture the complex interplay of demographic and clinical factors [18,19].

Incorporating artificial intelligence (AI) and its subset, machine learning (ML) into clinical care offers a critical and burgeoning opportunity to enhance patient-specific care by leveraging the staggering amount of demographic, radiographic, laboratory, and operative data collected during routine clinical visits [20−23]. Recently, Ames et al. [19] pioneered the use of unsupervised AI-based hierarchical clustering in spinal research to discern novel patient cohorts characterized by demographics, frailty indicators, radiographic metrics, and functional status.

The objective of the present study was to validate and extend the Ames-ISSG/ESSG classification using machine learning (ML)-based clustering analysis on a large, complex, multicenter, prospectively gathered adult spinal deformity (ASD) cohort. We concentrated on a comprehensive range of preoperative patient-reported outcomes (PROs), particularly physical and mental health metrics, frailty indices, and health status markers, hypothesizing that the interplay of mental health, frailty, and preoperative physical disability imposes a limit on the potential improvement of patients' PROs, particularly among heterogeneous elderly individuals. This investigation is the first to corroborate our clusters in an external dataset with long-term patient-reported and clinical outcomes.

Methods

Patient inclusion and exclusion criteria

The present study (NCT04194138) comprises a retrospective analysis performed on patients diagnosed with adult congenital, degenerative, idiopathic, or iatrogenic spinal deformities, with the primary deformity apex located between the cervicothoracic and thoracolumbar regions. All eligible participants satisfied at least one radiographic and one procedural inclusion criterion. Radiographic criteria encompassed pelvic incidence-lumbar lordosis (PI-LL) $\geq$ 25°, T1 Pelvic Angle (T1PA) $\geq$ 30°, Sagittal Vertical Axis (SVA) $\geq$ 15 cm, thoracic scoliosis $\geq$ 70°, thoracolumbar/lumbar (TL/L) scoliosis $\geq$ 50°, and global coronal malalignment $\geq$ 7 cm. Procedural criteria consisted of 3-column osteotomy (3CO) and spinal fusion with a minimum of 7 levels of instrumentation (See Supplementary Methods for Exclusion Criteria).

Training and validation datasets

We generated distinct patient phenotypes utilizing data from our prospectively collected dataset, which included 563 unique patients who underwent deformity correction between July 1, 2018, and October 1, 2022, across 18 spinal deformity centers. Subsequently, we validated the applicability and prognostic potential of our classification retrospectively using an external single-center, single-surgeon dataset. There was no overlap of patients between the training and validation cohorts; patients eligible for inclusion from the validation arm and present within the multicenter dataset were excluded from the validation analysis.

Within the single-center dataset, all patients underwent posterior spinal fusion (PSF) and had a minimum follow-up of 2 years. The primary outcome for the 4 clusters was the change in Scoliosis Research Society 22r Questionnaire scores from preoperative to 2-year time points, focusing specifically on the Function, Pain, Self-Image, Mental Health, and Satisfaction domains. Secondary outcome was intraoperative complication rate, readmission, and reoperation at 90-day, 1-year, 2-year, and 3-year timepoints.

Statistical analysis

The objective of this study was to develop preoperative phenotypes for ASD patients based on clinical and qualitative factors, rather than spinal alignment parameters which have previously been used in morphotyping [24]. We refined a comprehensive electronic health record dataset of over 500 potential data elements by incorporating aggregate health measures (eg, Charleson Comorbidity Index [CCI] for preoperative health status), selecting known health status indicators (eg, Edmonton Frailty Score), and excluding rare variables (<1% of cohort). (**List of abstracted variables are provided in Supplementary Methods**).

We employed unsupervised machine learning clustering. Patients were clustered based on their baseline patient-reported outcome measures, frailty, age, and preoperative health status (Supplementary Table 1). We utilized k-means clustering with Euclidean distance [25], which is a centroid-based clustering method that partitions a dataset into k clusters by minimizing the sum of squared distances within each cluster [26,27]. Briefly, each cluster is represented by its centroid, the arithmetic mean of the data points assigned to it. The process involves randomly selecting centroids for each cluster, assigning data points to the nearest centroid, recomputing centroids by averaging all points in a cluster, and repeating these steps until the sum of distances between data points and their corresponding centroids is minimized. This approach allows for the identification of distinct groups of, such that individuals within the same cluster exhibit greater similarity in their baseline characteristics compared to those in other clusters. Supplemental Figure 1 presents the silhouette coefficient for each patient. The silhouette coefficient, ranging from -1 to 1 (with 1 signifying an ideal fit of an observation to a phenotype). The mean silhouette score was 0.566, indicating an acceptable degree of cohesion and separation among the identified clusters.

After establishing clusters, we compared demographic, alignment, and clinical variables by cluster. Differences across clusters were tested using Brown-Forsythe and Welch's analysis of variance (ANOVA) with post-hoc Games-Howell pairwise comparisons for continuous variables and χ^2 [P(ChiSq)] tests for binary variables.

Next, to validate our clusters, we employed a random forests machine learning algorithm with two objectives: (1) to identify the most influential variables determining the phenotype a patient belongs to, and (2) to externally validate the applicability of the classification scheme. The random forest algorithm examined the association between a multitude of baseline variables and cluster assignment, and subsequently assigned patients within the validation cohort to the appropriate clusters garnered from the training cohort.

Within the validation cohort, to determine whether clusters were associated with differential trajectories we performed a mixed effects model to evaluate the changes in PROs over time. Cluster assignment and time (6 and 12 months) and cluster × time interaction was included as fixed effects. To evaluate clinical outcomes, the association between cluster and reoperation was examined using Kaplan-Meier estimators. Reoperation was then compared using a Cox model analysis adjusted for age, gender, total instrumented levels, and undergoing 3-column osteotomy.

Clustering was implemented using the Cluster package in R version 4.0 (R Project for Statistical Computing). Graphics were constructed using GraphPad Prism (v8.4.2) (GraphPad Software, La Jolla, CA, USA, www.graphpad.com). The remaining statistical analyses were performed in SAS version 9.4 (SAS Institute) with 2-tailed p<.05 to establish the statistical significance.

Results

Characterizing the multicenter prospective cohort for patient phenotype generation

Of the 563 ASD patients included in the analysis, the mean (Standard Error Mean[SEM]) age was 60.8 (0.64) years, the majority were female (379 [67.3%]), and the mean Edmonton Frailty score was 3.28 (0.1). 235 (41.7%) had prior thoracolumbar fusion. Among patients with prior thoracolumbar surgeries, 98 (41.7%) had ≥2 procedures (Table 1).

Demographics: age and frailty integration identifies four preoperative Phenotypes

Figure 1 illustrates the 4 phenotypes identified by the K-means algorithm, which were named using 3 axes [Age (**Y**oung vs. **O**ld), frailty (**F**rail vs. **R**esilient), and mental health (**C**ontent vs. **D**istressed)]. The "Old/Frail/Content" [OFC] cluster comprised of 156 patients (27.7%; green circles), "Old/Frail/Distressed" [OFD] cluster comprised of 187 patients (33.2%; purple squares), "Old/Resilient/Content"[ORC] cluster comprised of 153 patients (27.2%; blue triangles), and the "Young/ Resilient/Content" [YRC] cluster comprised of 67 patients (11.9%; red triangles).

Table 1
Characterization of demographics, baseline spinal alignment, and patient-reported outcomes in the training cohort of a prospective, multicenter clinical trial

	Total Cohort [N=563]
Demographics	
Age	60.75 (0.64)
Edmonton Frailty	3.28 (0.1)
Female Sex	379 (67.32)
Revision Surgery	199 (35.35)
Prior Thoracolumbar Fusion	235 (41.74)
1 Prior Procedure	135 (57.45)
2+ Prior Procedures	98 (41.7)
Spinal Alignment	
Max Cobb Angle	36.96 (1.03)
Sacral Slope[SS]	29.53 (0.56)
Pelvic Tilt[PT]	24.79 (0.49)
Pelvic Incidence[PI]	54.32 (0.57)
PI - LL	16.94 (1.01)
Lumbar Lordosis [LL]	37.37 (1.06)
Thoracic Kyphosis [TK]	-40.6 (0.99)
Cervical Lordosis [CL]	10.24 (0.73)
T1 Pelvic Angle [T1PA]	24.44 (0.61)
Cranial SVA to Hip [CrSVA-H]	36.92 (3.22)
SRS Schwab Classification − Coronal Curve	
N − No Coronal Curve	295 (52.4)
T − Thoracic Curve with Lumbar Curve < 30	7 (1.24)
L − Thoracolumbar Curve with Thoracic Curve < 30	141 (25.04)
D − Double Curve	120 (21.31)
SRS Schwab Classification − Pelvic Tilt	
0 − PT<20	220 (39.08)
+ − 20 < PT < 30	179 (31.79)
++ − PT > 30	164 (29.13)
SRS Schwab Classification − C7 Sagittal Vertical Axis	
0 − C7 SVA < 4 cm	254 (45.12)
+ − 4cm < C7SVA < 9.5 cm	150 (26.64)
++ − C7SVA > 9.5 cm	159 (28.24)
SRS Schwab Classification − PI − LL Modifier	
0 − PI-LL < 10	238 (42.27)
+ − 10 < PI-LL < 20	87 (15.45)
++ − PI-LL > 20	238 (42.27)
Baseline Patient Reported Outcome Scores	
B/L SRS Function	2.92 (0.04)
B/L SRS Pain	2.52 (0.04)
B/L SRS Image	2.44 (0.03)
B/L SRS Mental	3.57 (0.03)
B/L SRS Mean	2.86 (0.03)
B/L SRS Satisfaction	2.92 (0.04)
B/L SRS Total Mean	2.87 (0.03)
B/L ODI	43.64 (0.75)

Legend: A total of 563 adult patients with spinal deformity (ASD) were utilized to generate preoperative phenotypes. Data are presented as the mean, with the standard error of the mean (SEM) enclosed in parentheses.

Abbreviations: B/L refers to baseline or preoperative measure; SRS refers to the Scoliosis Research Society 22r questionnaire; ODI refers to the Oswestry Disability Index. All other abbreviations are specified within the text of the table; No. Prior Thoracolumbar Surgeries refers to the number of prior thoracolumbar procedures

Patients in the OFC Cluster had a mean (SEM) age of 67.9 (0.72) years, significantly older than OFD Cluster (61.4[0.89] years), ORC Cluster (62.3 [1.08] years), and YRC Cluster (38.9 [2.31] years) (all pairwise p<.0001). OFC and OFD clusters exhibited the highest Edmonton Frailty scores, with OFD being the frailest: OFC (3.76 [0.16]), OFD (4.72 [0.19]), ORC (2.10 [0.14]), and YRC (0.84 [0.15]) (all pairwise P < 0.0001; OFC vs. OFD: 0.0008). Additionally, the proportion of patients with prior thoracolumbar fusion showed significant differences among the 4 clusters. OFC and OFD Clusters had approximately 50% of patients with prior thoracolumbar procedures: OFC (47.4%), OFD (49.2%), ORC (36.6%), and YRC (19.4%) (p[ChiSq]<.0001) (Table 2).

Baseline alignment: confluence of global sagittal and coronal deformity differentiates phenotypes

The 4 unique phenotypes revealed specific spinal deformity patterns in both the coronal (SRS-Schwab Classifications T, L, or D) and global sagittal plane (SRS-Schwab C7 SVA Classifications + or ++). OFC Cluster had 69.9% (n=109) of patients with global sagittal deformity (p<.0001). In contrast, 70.2% (n=47) of YRC patients exhibited coronal deformity, with only 22.4% (n=15) presenting concomitant global sagittal deformity (Table 3). This observation aligns with YRC having the highest mean coronal Cobb Angle [OFC: 30.5(1.59) vs. OFD: 33.7(1.82) vs. ORC: 40.0(1.87) vs. YRC: 54.0(3.15), p<.0001 for YRC vs. all others] (Supplemental Table 1). Additionally, OFC had the highest T1PA, while YRC had the lowest [OFC: 29.0(1.03) vs. OFD: 24.9(0.94) vs. ORC: 24.6(1.16) vs. YRC: 11.9(1.74), p<.0001 for YRC vs. all; p<.05 for OFC vs. all] (Supplemental Table 1).

OFD and ORC exhibited similar alignment phenotypes to one another, with no significant differences in proportions of patients presenting with isolated global sagittal deformity (p[ChiSq]=.0812), isolated coronal deformity (p [ChiSq]=.0505), or mixed coronal and sagittal deformity (p [ChiSq]=.3273). The mean Coronal Cobb Angle, mean T1PA, and CrSVA-H were intermediate between OFC (which had the worst sagittal alignment) and YRC (which had the worst coronal alignment) (Supplemental Table 1).

Distinguishing phenotypes by functioning and mental health in the training Cohort

Numerous significant baseline pairwise differences were observed among patient-reported outcomes (Table 4). However, the four phenotypes were best distinguished by the intersection of their physical functioning/activity and mental health domains. YRC exhibited the best SRS-22r activity score (4.22±0.06) compared to OFC (2.48±0.04), OFD (2.37±0.04), and ORC (3.47±0.04) (p<.0001 for all pairwise comparisons) (Table 4).

The mental health score in OFD (2.71±0.04) was significantly worse than those in OFC (4.11±0.04, p<.0001), ORC (3.99±0.04, p<.0001), and YRC (3.71±0.08, p<.0001) demonstrating the worst physical function and mental health of the four clusters (Table 4).

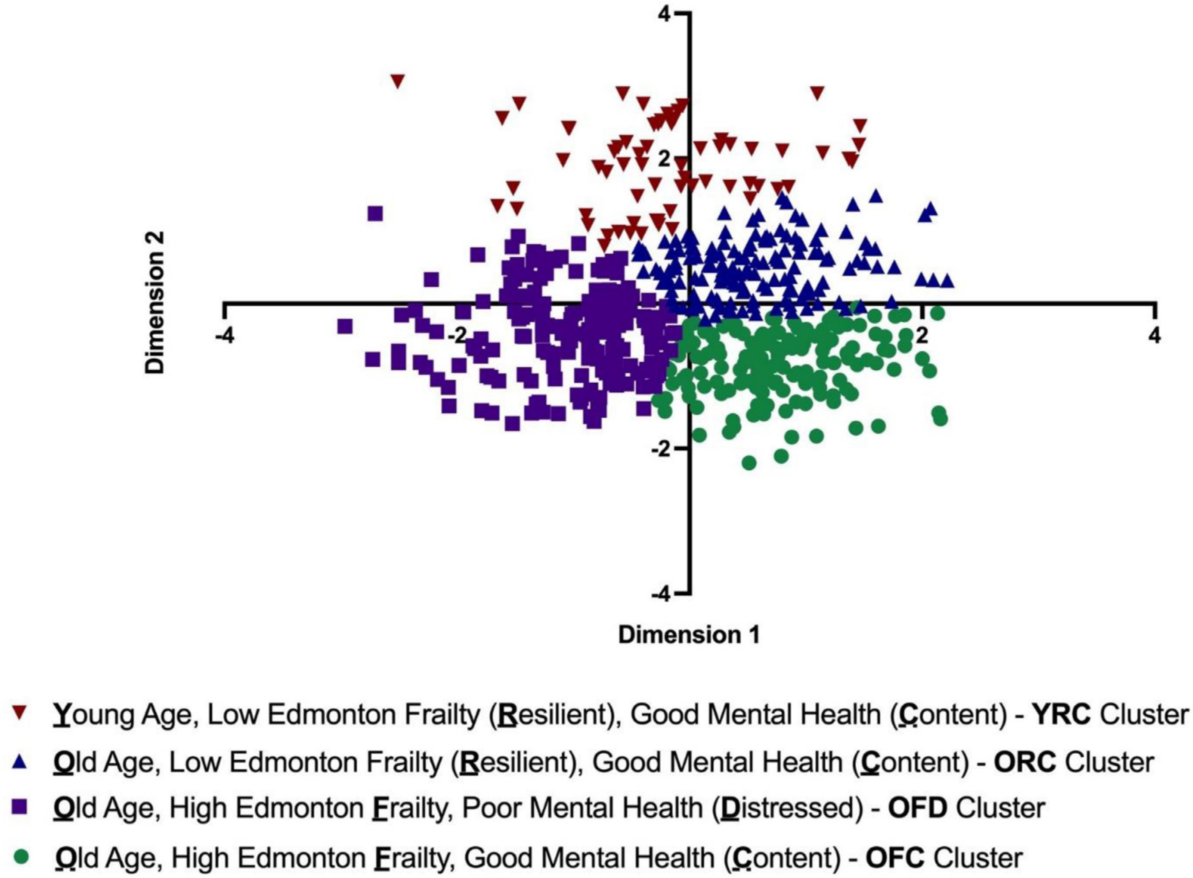

Fig. 1. Clustering of Adult Spinal Deformity (ASD) Patients Based on Preoperative Health Characteristics and Patient-Reported Outcomes. Four clusters were found using the k-means method with Euclidean distance. For the purpose of data visualization, the x- and y-axes are principal components patient reported outcome surveys and preoperative health measures including frailty, comorbidities, age, and surgical history.

This interaction is important, as OFC demonstrated similarly poor physical functioning as OFD (SRS-22r Activity: 2.48±0.04, p=.2115 vs. OFD) but significantly better mental health (SRS-22r Mental Health: 4.11±0.04, p<.0001) (Table 4).

Random forests regression classification for external validation in single-surgeon cohort

Our random forest regression model revealed that the top 5 patient-reported survey predictors, ranked in order of importance, were: VR-12 Physical Component Score (14.0), VR-12 Mental Component Score (12.8), SRS Mental Health (11.2), PROMIS Anxiety (11.1), and PROMIS Depression (9.57), highlighting the predominance of mental health-related PROs. Meanwhile, the top 5 variables other than patient-reported outcomes (PROs) ranked in order of importance were: Age (5.80), Edmonton Frailty (3.04), T1PA (2.61), PI-LL (2.05), and Coronal Cobb Angle (1.76) (Supplemental Figure 2).

The random forest algorithm analyzed the training cohort and assigned patients in a single-center validation cohort (n=242) to the four distinct clusters characterized as follows: OFC comprised of 156 patients (27.7%) (Supplemental Figure 4); OFD of 187 patients (33.2%) (Supplemental Figure 5); ORC of 153 patients (27.2%) (Supplemental Table 6); and YRC of 67 patients (11.9%) (Supplemental Table 7). The observed baseline pairwise differences among clusters in the training cohort were consistently identified in our single-center cohort, as presented in Supplemental Table 2 with representative case examples presented in Supplemental Figure 4-7.

Mixed effects model to compare postoperative improvements across patient clusters in the validation dataset

Within the validation cohort, all clusters exhibited improvement from preoperative to 2 years postop (p<.0001 for OFC, OFD, and ORC; p=.0832 for YRC). Patients in the OFC cluster experienced a significantly greater increase in SRS Function scores (improvement of 1.34 points, 95%CI, 1.01−1.67 points) compared to patients in the OFD cluster (improvement of 0.5 points, 95% CI, 0.245−0.755; p<.0001), the ORC cluster (improvement of 0.7 points, 95%CI, 0.415−0.985; p=.0045), and the YRC cluster (improvement of 0.24 points, 95%CI, -0.024 to 0.504; p<.0001).

Table 2
Demographic Profiles of the Four Preoperative Phenotypes Distinguished by the K-Means Clustering Algorithm

	Total Training Cohort [N=563]	OFC [N=156, **27.7%**]	OFD [N=187, **33.2%**]	ORC [N=153, **27.2%**]	YRC [N=67; **11.9%**]	OFC vs OFD	OFC vs ORC	OFC vs YRC	OFD vs ORC	OFD vs YRC	ORC vs YRC
Mean Silhouette Width	0.566	0.568	0.591	0.56	0.505	–	–	–	–	–	–
Demographics											
Age	60.75 (0.64)	67.86 (0.72)	61.38 (0.89)	62.29 (1.08)	38.96 (2.31)	**<0.0001**	**0.0001**	**<0.0001**	0.9154	**<0.0001**	**<0.0001**
Edmonton Frailty	3.28 (0.1)	3.76 (0.16)	4.72 (0.19)	2.1 (0.14)	0.84 (0.15)	**0.0008**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Female Sex	379 (67.32)	97 (62.18)	133 (71.12)	102 (66.67)	47 (70.15)	0.3736 *	–	–	–	–	–
Revision Surgery	199 (35.35)	62 (39.74)	81 (43.32)	45 (29.41)	11 (16.42)	**0.0002 ***	–	–	–	–	–
Prior Thoracolumbar Fusion	235 (41.74)	74 (47.44)	92 (49.2)	56 (36.6)	13 (19.4)	**<0.0001 ***	–	–	–	–	–
1 Prior Procedure	135 (57.45)	38 (51.35)	47 (51.09)	39 (69.64)	11 (84.62)	**0.0233 ***	–	–	–	–	–
2+ Prior Procedures	98 (41.7)	34 (45.95)	45 (48.91)	17 (30.36)	2 (15.38)		–	–	–	–	–

Legend: Data are presented as mean with standard error of the mean (SEM) for continuous variables or as the number of patients followed by the percentage of the cohort in parentheses for categorical variables. For continuous variables, P values are derived from the adjusted Brown-Forsythe and Welch's ANOVA, with Games-Howell post-hoc assessment for pairwise comparisons. For categorical variables [female sex, revision surgery, prior thoracolumbar fusion, number of prior thoracolumbar fusions], P values result from the Chi-Squared Test. The asterisk (*) indicates that the P value corresponds to the Chi-Squared Test and does not represent a pairwise comparison.

Table 3
Baseline alignment profiles of the four preoperative phenotypes distinguished by the k-means clustering algorithm

Schwab Classification Coronal Deformity	*Schwab Classification* C7 SVA Deformity	Deformity Indication	Total Cohort [N=563]	OFC [N=156, **27.7%**]	OFD [N=187, **33.2%**]	ORC [N=153, **27.2%**]	YRC [N=67; **11.9%**]	P Value [Indicated Row versus All Others]	P Value [OFD vs. ORC]
No Coronal Curve [N]	C7 SVA < 4cm [0]	Regional [TK or PI-LL]	129 (22.91%)	28 (17.95%)	52 (27.81%)	34 (22.22%)	15 (22.39%)	0.1899	0.2385
No Coronal Curve [N]	C7 SVA > 4cm [+ or ++]	Sagittal Alone	166 (29.48%)	65 (41.67%)	60 (32.09%)	36 (23.53%)	5 (7.46%)	**<0.0001**	0.0812
Any Coronal Curve [T, L, or D]	C7 SVA < 4cm [0]	Coronal Alone	125 (22.2%)	19 (12.18%)	30 (16.04%)	39 (25.49%)	37 (55.22%)	**<0.0001**	0.0505
Any Coronal Curve [T, L, or D]	C7 SVA > 4cm [+ or ++]	Mixed Coronal & Sagittal	143 (25.4%)	44 (28.21%)	45 (24.06%)	44 (28.76%)	10 (14.93%)	0.132	0.3273
-	C7 SVA > 4cm [+ or ++]	Any Global Sagittal	309 (54.88%)	109 (69.87%)	105 (56.15%)	80 (52.29%)	15 (22.39%)	**<0.0001**	n/a
Any Coronal Curve [T, L, or D]	-	Any Coronal	268 (47.6%)	63 (40.38%)	75 (40.11%)	83 (54.25%)	47 (70.15%)	**<0.0001**	n/a

Legend: Patients were grouped based on their coronal alignment (leftmost column) and global sagittal alignment (second column from the left), with the deformity indication being a combination of these two factors. To create statistically significant cohorts, the SRS-Schwab Classification was utilized. In the coronal plane, categories T, L, and D were combined into a single group, indicating the presence of any coronal curve. In the sagittal plane, + and ++ categories were merged to signify the presence of global sagittal deformity. Patients with any coronal curve and +/++ sagittal deformity were classified as "Mixed Coronal and Sagittal Deformity." Data are shown as the number of patients followed by the cohort percentage in parentheses. The P value corresponds to the results from the Chi-Squared Test, comparing the specified row against all other patients not in that row. The rightmost column displays the results of the Chi-Squared Test specifically comparing OFD and ORC.

Abbreviations: n/a indicates that the row was not evaluated statistically, a decision that was made *a priori.* For the SRS-Schwab Coronal Classification[N − No Coronal Curve, T − Thoracic Curve with Lumbar Curve < 30, L − Thoracolumbar Curve with Thoracic Curve < 30, D − Double Curve]. For the SRS-Schwab Sagittal Classification[0 − C7 SVA < 4cm, + − 4cm < C7SVA < 9.5cm, ++ − C7SVA > 9.5cm]

The relative magnitude of improvement experienced by the four clusters was also evident in the SRS Pain domain. Patients in the OFC cluster improved significantly more than those in all other clusters. Notably, patients in the OFD cluster, characterized by poor preoperative physical functioning and mental health, consistently improved the least, contrasting sharply with patients in the OFC cluster, who had comparable preoperative physical functioning and pain scores (Table 5).

Interestingly, all clusters demonstrated nearly identical improvement in SRS Mental Health scores, with no significant difference in the magnitude of change for all pairwise comparisons (OFC: 0.37 points vs. OFD: 0.31 points vs. ORC: 0.46 points vs. YRC: 0.3 points, p>.69 for all pairs) (Table 5).

Clinical outcomes across patient clusters in the validation dataset

To assess the impact of patient phenotype on reoperation, we performed a multivariable Cox regression analysis, adjusting for age, gender, total instrumented levels, and three-column osteotomy procedures. Kaplan-Meier survival curves revealed that the reoperation-free survival rates significantly differed among YRC, ORC, OFC, and OFD, listed from best to worst outcomes (log-rank test, $\chi 2$=7.456 and p=.0063) (Fig. 2). After adjusting for confounding factors, the OFD cohort showed significantly worse two-year reoperation outcomes (HR: 3.303 [95%CI: 1.085−8.390]) compared to the OFC cohort, whereas ORC (HR: 0.6517 [95%CI: 0.088−3.436]) and YRC (HR: 0.3862 [95%CI: 0.0191−2.799]) demonstrated statistically similar results (Table 6).

Discussion

Our utilization of ML in two separate ASD cohorts reinforces the capability of unsupervised clustering to discern subtle patterns, categorize patients, reveal previously unconsidered variables, and predict patient trajectories. Data-driven AI/ML clustering approaches can incorporate a plethora of variables into classification schemes, enabling granular patient population segmentation. Given the rapid evolution of AI/ML, continuous comparison with existing frameworks remains crucial.

Review of AI/ML applications in spinal deformity: how do we cluster ASD patients preoperatively?

Utilizing a systematic review algorithm, we identified 226 unique titles and 66 full-text articles, ultimately distilling them down to four studies (Supplemental Methods; Supplemental Table/Figure 3). Among the included studies, one study [19] suggested a preoperative qualitative classification for ASD patients, whereas the remaining studies [24,28,29] aimed to characterize distinct spine shapes by proposing patient clusters solely based on spinal alignment measures.

Table 4
Comparison of pain, functioning, and mental health measures at baseline across identified clusters

	Total Training Cohort [N=563]	OFC [N=156, **27.7%**]	OFD [N=187, **33.2%**]	ORC [N=153, **27.2%**]	YRC [N=67; **11.9%**]	OFC vs OFD	OFC vs ORC	OFC vs YRC	OFD vs ORC	OFD vs YRC	ORC vs YRC
NRS Back and Leg Pain											
Back	6.87 (0.1)	7.36 (0.18)	7.72 (0.14)	6.49 (0.17)	4.25 (0.31)	0.3923	**0.0028**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Leg	4.2 (0.14)	4.74 (0.26)	5.41 (0.23)	3.43 (0.26)	1.31 (0.26)	0.2175	**0.0024**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Scoliosis Research Society-22 (SRS-22r)											
Activity	2.92 (0.04)	2.48 (0.04)	2.37 (0.04)	3.47 (0.04)	4.22 (0.06)	0.2115	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Pain	2.52 (0.04)	2.17 (0.05)	2.06 (0.05)	2.94 (0.05)	3.67 (0.08)	0.4056	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Appearance	2.44 (0.03)	2.41 (0.05)	1.95 (0.04)	2.87 (0.05)	2.85 (0.08)	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	0.9966
Mental Health	3.57 (0.03)	4.11 (0.04)	2.71 (0.04)	3.99 (0.04)	3.71 (0.08)	**<0.0001**	0.1486	**<0.0001**	**<0.0001**	**<0.0001**	**0.0114**
Satisfaction	2.92 (0.04)	2.81 (0.08)	2.78 (0.08)	3.03 (0.08)	3.28 (0.12)	0.9935	0.2117	**0.0074**	0.1227	**0.0037**	0.3098
Total Mean	2.87 (0.03)	2.79 (0.03)	2.32 (0.03)	3.29 (0.03)	3.58 (0.05)	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Veterans RAND 12 Item Health Survey (VR-12)											
Physical Component Score	29.34 (0.46)	19.62 (0.47)	26.64 (0.51)	34.44 (0.56)	47.88 (0.85)	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Mental Component Score	48.25 (0.53)	55.5 (0.71)	35.27 (0.65)	55.9 (0.61)	50.09 (1.08)	**<0.0001**	0.9738	**0.0003**	**<0.0001**	**<0.0001**	**<0.0001**
Oswestry Disability Index (ODI)											
ODI	43.64 (0.75)	52.98 (0.93)	54.34 (0.92)	32.99 (0.88)	15.08 (1.03)	0.7263	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**

Legend: This table presents the mean values and standard errors for NRS Back and Leg Pain, SRS-22r, VR-12, and ODI scores for the total cohort and each identified cluster. P-values are provided for pairwise comparisons between clusters.

Abbreviations: NRS: Numeric Rating Scale; SRS-22r: Scoliosis Research Society-22 Revised Questionnaire; VR-12: Veterans RAND 12-Item Health Survey; ODI: Oswestry Disability Index.

Table 5
Mean Change in SRS-22r Domains Over Time in a Single-Center, Single-Surgeon Validation Cohort

	Are the *changes* in PROs significantly different **within** clusters?				Are the *changes* in PROs significantly different between clusters?					
	OFC [N=42, **17.35%**]	OFD [N=71 [**29.33%**]	ORC [N=60 **24.79%**]	YRC [N=69 **28.51%**]	OFC vs OFD	OFC vs ORC	OFC vs YRC	OFD vs ORC	OFD vs YRC	ORC vs YRC
Baseline SRS Function	**2.45 (0.06)**	**2.54 (0.04)**	**3.6 (0.04)**	**4.34 (0.04)**						
Baseline to One-Year	0.98 (<0.0001)	0.15 (0.3506)	0.52 (<0.0001)	-0.05 (0.8965)	<0.0001	0.0692	<0.0001	0.1098	0.5766	0.0042
Baseline to Two-Year	1.34 (<0.0001)	0.5 (<0.0001)	0.7 (<0.0001)	0.24 (0.0832)	<0.0001	0.0045	<0.0001	0.6097	0.346	0.0312
One-Year to Two-Year	0.36 (0.0279)	0.35 (0.0037)	0.18 (0.2992)	0.29 (0.027)	>0.9999	0.766	0.9798	0.7236	0.9806	0.9099
Baseline SRS Pain	**2.26 (0.07)**	**2.2 (0.04)**	**3.18 (0.05)**	**3.93 (0.04)**						
Baseline to One-Year	0.93 (<0.0001)	0.18 (0.3003)	0.77 (<0.0001)	0.31 (0.037)	0.0016	0.8672	0.014	0.0083	0.8792	0.0671
Baseline to Two-Year	1.46 (<0.0001)	0.87 (<0.0001)	0.8 (<0.0001)	0.52 (0.0001)	0.0193	0.0105	<0.0001	0.9806	0.1921	0.4327
One-Year to Two-Year	0.53 (0.0022)	0.69 (<0.0001)	0.03 (0.9735)	0.21 (0.2174)	0.8512	0.082	0.3879	0.0024	0.0341	0.7652
Baseline SRS Image	**2.44 (0.07)**	**2.03 (0.04)**	**2.81 (0.04)**	**3.09 (0.05)**						
Baseline to One-Year	1.45 (<0.0001)	0.28 (0.026)	1.36 (<0.0001)	1.19 (<0.0001)	<0.0001	0.9615	0.468	<0.0001	<0.0001	0.7297
Baseline to Two-Year	1.42 (<0.0001)	1.31 (<0.0001)	1.39 (<0.0001)	1.28 (<0.0001)	0.924	0.9985	0.8615	0.9602	0.9974	0.9085
One-Year to Two-Year	-0.03 (0.9747)	1.03 (<0.0001)	0.03 (0.9665)	0.09 (0.6994)	<0.0001	0.988	0.9074	<0.0001	<0.0001	0.9833
Baseline SRS Mental	**3.99 (0.04)**	**2.76 (0.04)**	**3.83 (0.05)**	**3.99 (0.05)**						
Baseline to One-Year	0.16 (0.393)	0.13 (0.358)	0.19 (0.1739)	0.23 (0.0512)	0.9974	0.9977	0.9704	0.9748	0.8843	0.9926
Baseline to Two-Year	0.37 (0.0074)	0.31 (0.0033)	0.46 (<0.0001)	0.3 (0.0067)	0.9802	0.945	0.9704	0.7187	0.9999	0.6868
One-Year to Two-Year	0.21 (0.2012)	0.18 (0.1408)	0.27 (0.0301)	0.07 (0.7565)	0.9974	0.9826	0.8096	0.9217	0.8522	0.513
Baseline SRS Satisfaction	**2.86 (0.1)**	**2.55 (0.07)**	**3.08 (0.07)**	**3.33 (0.06)**						
Baseline to One-Year	1.41 (<0.0001)	0.16 (0.4486)	1.35 (<0.0001)	1.17 (<0.0001)	<0.0001	0.9934	0.6932	<0.0001	<0.0001	0.8089
Baseline to Two-Year	1.3 (<0.0001)	0.98 (<0.0001)	1.19 (<0.0001)	1.28 (<0.0001)	0.4535	0.9619	0.9997	0.7159	0.3967	0.9703
One-Year to Two-Year	-0.11 (0.7961)	0.82 (<0.0001)	-0.16 (0.5263)	0.11 (0.7015)	0.0002	0.9962	0.7474	<0.0001	0.0016	0.5404

Legend. Mean changes from baseline were estimated with a generalized linear mixed-effects model for repeated measures for change in SRS-22r Function, Pain, Self-Image, Mental Health, and Satisfaction domains, with fixed effects for baseline SRS-score, age, sex, time, and with random effects for clusters.

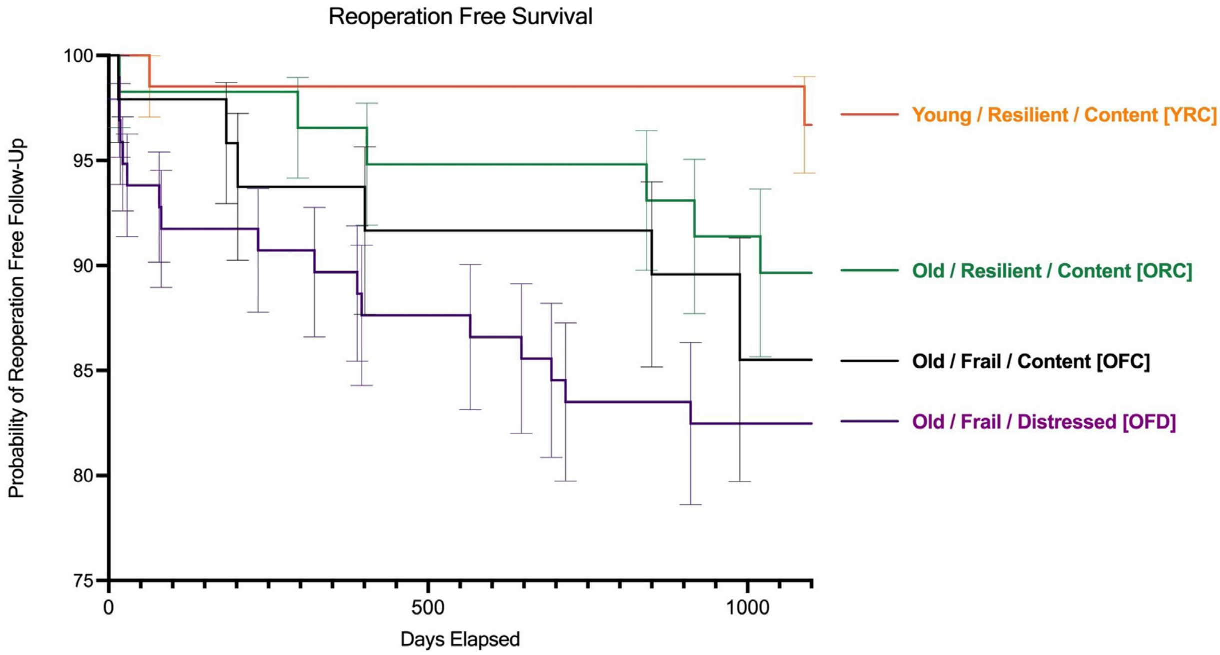

Fig. 2. Kaplan-Meier Estimated Reoperation Free Survival Curves

We qualitatively compared our study to the only other analysis assessing qualitative, preop clustering of ASD patients by Ames and colleagues [19]. Methodologically, Ames and colleagues clustered "patient related factors" [age, sex, height, weight, and number of previous spine surgeries], "radiographic factors" [coronal and sagittal measures], and PROs[ODI, SRS-22r, and SF36v2], but it is unclear exactly which variables, and scores were input into the clustering algorithm. In contrast, we clustered variables listed in the Supplemental Methods, which include the subdomains of all PROs[SRS22r, ODI, PROMIS, VR12, NRS], frailty [Edmonton Frailty domains], and patient characteristics [Age, Number of comorbidities, Charleson comorbidity, functional status, and spine surgical history] but importantly excluded alignment measures.

When comparing our clusters, we find that Ames' "YC: Young Coronal" cohort is nearly identical to our YRC cluster. Operative treatment for the YRC cohort is driven by pronounced coronal plane deformity and concerns about further curve progression [30], while robust compensatory mechanisms mitigate sagittal deformity. Uniquely, motivation for operative care within YRC cohort is a function of diminished self-image perception, mild-moderate back pain (less severe than in older cohorts), and partial functional impairment− the confluence of which affects their quality of life [31,32]. Consistently, in this sub-population, SRS Self-Image domain has been reported to demonstrate the strongest correlation with patient satisfaction − which is distinct from the ORC, OFC, ORD cohorts [33]. Moreover, in YRC patients, PROs might be an insufficient marker of operative success. That is, it could be impossible to achieve MCID thresholds due to the ceiling of the instruments deployed, the fact that worse baseline health scores predict more significant quantitative PROM improvement [11], and an inherent limit to quantitative improvement given a high functional baseline.

Similar to Ames' classification, the "OPrim: Old Primary" cluster resembles our ORC cluster which have similar age (61.0[19] vs. 62.3), moderate sagittal malalignment (66.8[19] vs. 47.8), and similar coronal Cobb angle (36.6[19] vs. 40.03). However, our ORC cluster had superior SRS-total score at baseline compared to the OPrim cluster [19] (3.29 vs. 2.78[19]). Consequently, we theorize that a proportion of elderly patients undergoing a primary procedure (OPrim) are marked by increased frailty, poor physical function, and poor mental health that are captured within the OPrim cluster − but are prone to limited PRO improvement and higher complication rates.

The most notable divergence between the two analyses lies in our hypothesis that older, frail patients, including those undergoing revision, constitute a heterogeneous group vulnerable to distinct postoperative trajectories. Our examination of changes in PROs (Table 5), KM curves (Fig. 2), and analyses of complications and reoperations (Table 6) collectively suggest that OFC and OFD differ significantly. The primary factors differentiating these patients are baseline mental health (VR12 MCS, SRS-Mental Health) and frailty.

The discrepancy between our cluster analysis and Ames and colleagues', despite similar overarching approaches, could stem from the different tools used to assess mental health, pain, and frailty. For instance, PROMIS allows for a more refined evaluation of functional limitations by incorporating pain interference, anxiety, and depression. It correlates strongly with ODI [34] and SRS-22r [35], presenting better ceiling and floor effects than SRS-22r [36]. The

Table 6
Clinical outcomes across clusters in a single-center, single-surgeon validation cohort

	OFC [N=42, 17.35%]	OFD [N=71 [29.33%]	ORC [N=60 24.79%]	YRC [N=69 28.51%]	Unadjusted P Value	Cox proportional hazards model [Hazard Ratio (95% Confidence Interval)] OFC	OFD	ORC	YRC
Any Intraop Complications	19 (45.24)	42 (59.15)	19 (31.67)	12 (17.39)	<0.0001				
Intraop Dural Tear	14 (33.33)	22 (30.99)	8 (13.33)	5 (7.25)	not evaluated *				
Intraop Neuromonitoring Change	3 (7.14)	13 (18.31)	6 (10)	4 (5.8)	not evaluated *				
Postop Inpatient Complication	8 (19.05)	8 (11.27)	16 (26.67)	9 (13.04)	not evaluated *				
Readmissions									
90-Day Readmissions	3 (7.14)	7 (9.86)	1 (1.67)	1 (1.45)	0.065				
One-Year Readmissions	5 (11.9)	9 (12.68)	2 (3.33)	1 (1.45)	0.0229				
Two-Year Readmissions	6 (14.29)	13 (18.31)	6 (10)	1 (1.45)	0.0115				
Three Year Readmissions	10 (23.81)	17 (23.94)	7 (11.67)	1 (1.45)	0.0005				
Reoperations									
90-Day Reoperations	1 (2.38)	6 (8.45)	1 (1.67)	1 (1.45)	0.0902	1.00 [Reference]	6.497 (1.145 to 122.4)	1.61 (0.062 to 41.44)	2.168 (0.082 to 57.06)
One-Year Reoperations	2 (4.76)	6 (8.45)	1 (1.67)	1 (1.45)	0.1352	1.00 [Reference]	3.189 (1.062 to 21.75)	0.7401 (0.0338 to 7.958)	0.9184 (0.040 to 10.34)
Two-Year Reoperations	4 (9.52)	11 (15.49)	2 (3.33)	1 (1.45)	0.0073	1.00 [Reference]	3.303 (1.085 to 8.390)	0.6517 (0.088 to 3.436)	0.3862 (0.0191 to 2.799)
Three Year Reoperations	7 (16.67)	15 (21.13)	7 (11.67)	1 (1.45)	0.0094	1.00 [Reference]	1.481 (0.622 to 3.911)	0.9187 (0.287 to 2.848)	0.166 (0.0087 to 0.9948)

capacity of PROMIS and VR12 to capture essential aspects of pain and mental health may have contributed to the significant divergence between OFC and OFD clusters in our analysis, while the prior analysis considered elderly, revision patients more homogeneously [37,38]. As predictive analytics advances, it is crucial to emphasize the validity, robustness, and relevance of input variables.

Confluence of patient qualitative phenotype and deformity type drive Complications and PROs

Apart from the study by Ames et al. [19], other included analyses utilize ML/AI algorithms to cluster sagittal morphology of ASD patients, using alignment measures that were purposefully omitted from our clustering algorithms. Durand and colleagues [24], for example, reviewed 915 ASD preoperative lateral radiographs and employed an artificial neural network to discern 6 morphologies characterizing sagittal profiles among ASD patients. Qualitatively, these phenotypes consisted of varying degrees of positive sagittal balance, lower extremity compensation (pelvic retroversion), and lumbo-pelvic alignment. Intriguingly, all clusters displayed comparable improvements in PROs over a 2-year follow-up but differed in clinical outcomes such as PJF incidence, ranging from 5% to 18% across clusters − a figure comparable to the two-year reoperation range observed across our clusters (1.45%−21.13%) [24]. More recently, Lafage and colleagues proposed a simplified classification of preoperative alignment among ASD patients, introducing 4 cohorts: hyper-thoracic kyphosis, moderate sagittal imbalance, severe sagittal imbalance, and severe coronal imbalance [28]. Nonetheless, this study only reported a 30-day follow-up, which prevents drawing conclusions on outcomes over an extended period.

Combined with the findings presented here, the literature suggests that a patient's expected complication rate and PRO improvement can be determined by mapping their qualitative phenotype, as defined in our study, and their spinal deformity phenotype. We suggest that the main determinant of patients' PROs is the qualitative phenotype, while the alignment phenotype plays a lesser role. In contrast, the primary factor influencing patients' clinical outcomes such as PJK/PJF is likely their alignment phenotype, with the qualitative phenotype contributing to a smaller extent.

Limitations

This study has several limitations beyond those inherent to retrospective analyses. Firstly, the qualitative classification is confined to operative ASD patients treated by members of our study group, excluding nonoperative patients, cervical spine deformity cases, and those from different clinical sites, which could reveal a broader range of patient phenotypes. Similarly, although the clusters encompass most patients presenting at an ASD center, it is possible that certain phenotypes, such as young, resilient, and distressed (YRD) patients, exist but were not adequately represented in our cohorts to be identified. Secondly, the intentional exclusion of alignment measures from the clustering algorithm precludes a comprehensive evaluation of a patient's presentation, which depends on both their physical and mental disposition and their pathology. Combining the clusters presented here and those in Ames [19] with those from Lafage [28] or Durand [24] could produce a more comprehensive classification scheme. Thirdly, due to the immaturity of our training cohort, we were unable to track long-term outcomes concerning these patient clusters. Lastly, to mitigate the "black box" nature of AI/ML models, future works will concentrate on enhancing the accessibility of phenotyping patients using actionable pathways.

Conclusion

In our multicenter, prospective cohort study involving 563 ASD patients, machine learning clustering discerned four distinct preoperative phenotypes— "Old/Frail/Content" (OFC), "Old/Frail/Distressed" (OFD), "Old/Resilient/Content" (ORC), and "Young/Resilient/Content" (YRC)—predicated on age, frailty, physical functioning, and mental health. When corroborated in a single-surgeon validation cohort, all phenotypes showed postoperative improvements in PROs over 2-years, but the OFC cluster exhibited significantly greater improvements in SRS Function and Pain scores, whereas SRS Mental Health improved comparably across all clusters. Clinically, the OFD cluster exhibited worse reoperation outcomes compared to all other clusters, substantiating that a patient's preoperative qualitative phenotype influences patients' capacity to improve their PROs following ASD correction.

Declaration of Competing Interest

One or more of the authors declare financial or professional relationships on ICMJE-TSJ disclosure forms.

CRediT authorship contribution statement

Sarthak Mohanty: Conceptualization, Data curation, Formal analysis, Methodology, Writing − original draft, Investigation, Writing − review & editing. **Fthimnir M. Hassan:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing − original draft, Writing − review & editing. **Lawrence G. Lenke:** Conceptualization, Data curation, Investigation, Supervision, Validation, Writing − review & editing, Project administration. **Erik Lewerenz:** Conceptualization, Data curation, Writing − review & editing. **Peter G. Passias:** Conceptualization, Data curation, Validation, Writing − review & editing. **Eric O. Klineberg:** Conceptualization, Data curation, Validation, Writing − review & editing. **Virginie Lafage:** Conceptualization, Data curation, Validation, Writing − review & editing. **Justin S. Smith:** Conceptualization, Data curation, Validation, Writing − review & editing. **D. Kojo Hamilton:** Conceptualization, Data curation, Validation, Writing − review & editing. **Jeffrey L. Gum:** Conceptualization, Data curation, Validation, Writing − review & editing. **Renaud Lafage:** Conceptualization, Data curation, Validation, Writing − review & editing. **Jeffrey Mullin:** Conceptualization, Data curation, Validation, Writing − review & editing. **Bassel Diebo:** Conceptualization, Data curation, Validation, Writing − review & editing. **Thomas J. Buell:** Conceptualization, Data curation, Validation, Writing − review & editing. **Han Jo Kim:** Conceptualization, Data curation, Validation, Writing − review & editing. **Khalid Kebaish:** Conceptualization, Data curation, Validation, Writing − review & editing. **Robert Eastlack:** Conceptualization, Data curation, Visualization, Writing − review & editing. **Alan H. Daniels:** Conceptualization, Data curation, Validation, Writing − review & editing. **Gregory Mundis:** Conceptualization, Data curation, Validation, Writing − review & editing. **Richard Hostin:** Conceptualization, Data curation, Validation, Writing − review & editing. **Themistocles S. Protopsaltis:** Conceptualization, Data curation, Validation, Writing − review & editing. **Robert A. Hart:** Conceptualization, Data curation, Validation, Writing − review & editing. **Munish Gupta:** Conceptualization, Data curation, Validation, Writing − review & editing. **Frank J. Schwab:** Conceptualization, Data curation, Validation, Writing − review & editing. **Christopher I. Shaffrey:** Conceptualization, Data curation, Validation, Writing − review & editing. **Christopher P. Ames:** Conceptualization, Data curation, Validation, Writing − review & editing. **Douglas Burton:** Conceptualization, Data curation, Validation, Writing − review & editing. **Shay Bess:** Conceptualization, Data curation, Project administration, Supervision, Validation, Writing − review & editing.

Acknowledgments

International Spine Study Group (ISSG) is an independent, multi-center study group devoted to performing the highest quality research to evaluate the clinical and radiographic outcomes associated with treatment for adult spinal deformity. All funding for ISSG study efforts comes from the International Spine Study Group Foundation (ISSGF) which is a non-profit, independent, private foundation, 501 (3) (c) established 2/1/2010.

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.spinee.2024.02.010.

References

[1] Bess S, Line B, Fu K-M, McCarthy I, Lafage V, Schwab F, et al. The health impact of symptomatic adult spinal deformity: comparison of deformity types to United States Population norms and chronic diseases. Spine (Phila Pa 1976) 2016;41(3):224–33. https://doi.org/10.1097/BRS.0000000000001202.

[2] Schwab F, Dubey A, Gamez L, El Fegoun A-B, Hwang K, Pagala M, et al. Adult scoliosis: prevalence, SF-36, and nutritional parameters in an elderly volunteer population. Spine (Phila Pa 1976) 2005;30 (9):1082–5. https://doi.org/10.1097/01.brs.0000160842.43482.cd.

[3] Ames CP, Scheer JK, Lafage V, Smith JS, Bess S, Berven SH, et al. Adult spinal deformity: epidemiology, health impact, evaluation, and management. Spine Deform 2016;4(4):310–22. https://doi.org/10.1016/j.jspd.2015.12.009.

[4] Pellisé F, Vila-Casademunt A, Ferrer M, Domingo-Sabat M, Bago J, Perez-Grueso FJS, et al. Impact on health related quality of life of adult spinal deformity (ASD) compared with other chronic conditions. Eur spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc 2015;24(1):3–11. https://doi.org/10.1007/s00586-014-3542-1.

[5] Bridwell KH, Baldus C, Berven S, Edwards II C, Glassman S, Hamill C, et al. Changes in radiographic and clinical outcomes with primary treatment adult spinal deformity surgeries from two years to three- to five-years follow-up. Spine (Phila Pa 1976) 2010;35(20):1849–54. https://doi.org/10.1097/BRS.0b013e3181efa06a.

[6] Smith JS, Shaffrey CI, Glassman SD, Berven SH, Schwab FJ, Hamill CL, et al. Risk-benefit assessment of surgery for adult scoliosis: an analysis based on patient age. Spine (Phila Pa 1976) 2011;36 (10):817–24. https://doi.org/10.1097/BRS.0b013e3181e21783.

[7] Kelly MP, Lurie JD, Yanik EL, Shaffrey CI, Baldus CR, Boachie-Adjei O, et al. Operative versus nonoperative treatment for adult symptomatic lumbar scoliosis. J Bone Joint Surg Am 2019;101 (4):338–52. https://doi.org/10.2106/JBJS.18.00483.

[8] Smith JS, Shaffrey CI, Berven S, Glassman S, Hamill C, Horton W. Operative versus nonoperative treatment of leg pain in adults with scoliosis: a retrospective review of a prospective multicenter database with two-year follow-up. Spine (Phila Pa 1976) 2009;34:1693–8. https://doi.org/10.1097/BRS.0b013e3181ac5fcd.

[9] 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;34(20):2171–8. https://doi.org/10.1097/BRS.0b013e3181a8fdc8.

[10] Whitebird RR, Solberg LI, Norton CK, Ziegenfuss JY, Asche SE, Grossman ES. What outcomes matter to patients after joint or spine surgery? J Patient Cent Res Rev 2020;7(2):157–64. https://doi.org/10.17294/2330-0698.1738.

[11] Wondra JP, Kelly MP, Yanik EL, Greenberg JK, Smith JS, Bess S, et al. Patient-reported outcome measure clustering after surgery for adult symptomatic lumbar scoliosis. J Neurosurg Spine 2022;37 (1):80–91. https://doi.org/10.3171/2021.11.SPINE21949.

[12] Moal B, Lafage V, Smith JS. Clinical improvement through surgery for adult spinal deformity: what can be expected and who is likely to benefit most? Spine Deform 2015;3(6):566–74.

[13] Pierce KE, Passias PG, Alas H. Does patient frailty status influence recovery following spinal fusion for adult spinal deformity? An analysis of patients with 3-year follow-up. Spine (Phila Pa 1976) 2020;45 (7):E397–405.

[14] Terran J, Schwab F, Shaffrey CI, Smith JS, Devos P, Ames CP, et al. The SRS-Schwab adult spinal deformity classification: assessment and clinical correlations based on a prospective operative and nonoperative cohort. Neurosurgery 2013;73(4):559–68. https://doi.org/10.1227/NEU.0000000000000012.

[15] Smith JS, Klineberg E, Schwab F, Shaffrey CI, Moal B, Ames CP, et al. Change in classification grade by the SRS-Schwab Adult Spinal Deformity Classification predicts impact on health-related quality of life measures: prospective analysis of operative and nonoperative treatment. Spine (Phila Pa 1976) 2013;38(19):1663–71. https://doi.org/10.1097/BRS.0b013e31829ec563.

[16] Schwab F, Ungar B, Blondel B, Buchowski J, Coe J, Deinlein D, et al. Scoliosis Research Society-Schwab adult spinal deformity classification: a validation study. Spine (Phila Pa 1976) 2012;37 (12):1077–82. https://doi.org/10.1097/BRS.0b013e31823e15e2.

[17] Lenke LG, Betz RR, Harms J, Bridwell KH, Clements DH, Lowe TG, et al. Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis. J Bone Joint Surg Am 2001;83 (8):1169–81.

[18] Oh T, Scheer JK, Smith JS, Hostin R, Robinson C, Gum JL, et al. Potential of predictive computer models for preoperative patient selection to enhance overall quality-adjusted life years gained at 2-year follow-up: a simulation in 234 patients with adult spinal deformity. Neurosurg Focus 2017;43(6):E2. https://doi.org/10.3171/2017.9.FOCUS17494.

[19] Ames CP, Smith JS, Pellisé F, Kelly M, Alanay A, Acaroglu E, et al. Artificial intelligence based hierarchical clustering of patient types and intervention categories in adult spinal deformity surgery: towards a new classification scheme that predicts quality and value. Spine (Phila Pa 1976) 2019;44(13):915–26. https://doi.org/10.1097/BRS.0000000000002974.

[20] Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med 2019;380(14):1347–58. https://doi.org/10.1056/NEJMra1814259.

[21] Meiring C, Dixit A, Harris S, MacCallum NS, Brealey DA, Watkinson PJ, et al. Optimal intensive care outcome prediction over time using machine learning. PLoS One 2018;13(11):e0206862.

[22] Kwon J, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. J Am Heart Assoc 2018;7(13):e008678. https://doi.org/10.1161/JAHA.118.008678.

[23] Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018;24(11):1716–20. https://doi.org/10.1038/s41591-018-0213-5.

[24] Durand WM, Lafage R, Hamilton DK, Passias PG, Kim HJ, Protopsaltis T, et al. Artificial intelligence clustering of adult spinal deformity sagittal plane morphology predicts surgical characteristics, alignment, and outcomes. Eur spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc 2021;30(8):2157–66. https://doi.org/10.1007/s00586-021-06799-z.

[25] Singh A, Yadav A, Rana A. K-means with Three different Distance Metrics. Int J Comput Appl 2013;67(10):14–7.

[26] Hartigan JA, Wong MA. A k-means clustering algorithm. Appl Stat 1979;28(1):100–8.

[27] Xu H, Intrator O, Culakova E, Bowblis JR. Changing landscape of nursing homes serving residents with dementia and mental illnesses. Health Serv Res 2022;57(3):505–14. https://doi.org/10.1111/1475-6773.13908.

[28] Lafage R, Fourman MS, Smith JS, Bess S, Shaffrey CI, Kim HJ, et al. Can unsupervised cluster analysis identify patterns of complex adult spinal deformity with distinct perioperative outcomes? J Neurosurg Spine 2023;38(5):547–57. https://doi.org/10.3171/2023.1.SPINE221095.

[29] Kim HJ, Virk S, Elysee J, Passias P, Ames C, Shaffrey CI, et al. The morphology of cervical deformities: a two-step cluster analysis to identify cervical deformity patterns. J Neurosurg Spine 2019:1–7 Published online November. https://doi.org/10.3171/2019.9.SPINE19730.

[30] Kim HJ, Yang JH, Chang D-G, Suk S-I, Suh SW, Song K-S, et al. Adult spinal deformity: current concepts and decision-making strategies for management. Asian Spine J 2020;14(6):886–97. https://doi.org/10.31616/asj.2020.0568.

[31] Pizones J, Moreno-Manzanaro L, Vila-Casademunt A, Fernandez-Baillo N, Sanchez-Marquez J, Talavera G, et al. Adult congenital spine deformity: clinical features and motivations for surgical treatment. Int J spine Surg 2021;15(6):1238–45. https://doi.org/10.14444/8157.

[32] Bess S, Boachie-Adjei O, Burton D, Cunningham M, Shaffrey C, Shelokov A, et al. Pain and disability determine treatment modality

for older patients with adult scoliosis, while deformity guides treatment for younger patients. Spine (Phila Pa 1976) 2009;34(20):2186–90. https://doi.org/10.1097/BRS.0b013e3181b05146.

[33] Gum JL, Shasti M, Yeramaneni S, Carreon LY, Hostin RA, Kelly MP, et al. Improvement in SRS-22R self-image correlate most with patient satisfaction after 3-column osteotomy. Spine (Phila Pa 1976) 2021;46(12):822–7. https://doi.org/10.1097/BRS.0000000000003897.

[34] Stekas ND, Johnson B, Jevotovsky D. PROMIS is superior to established outcome measures in capturing disability resulting from sagittal malalignment in patients with back pain. Spine Deform 2020;8(3):499–505.

[35] Ibaseta A, Rahman R, Skolasky RL, Reidler JS, Kebaish KM, Neuman BJ. SRS-22r legacy scores can be accurately translated to PROMIS scores in adult spinal deformity patients. Spine J 2020;20(2):234–40.

[36] Bernstein DN, Papuga MO, Sanders JO, Rubery PT, Menga EN, Mesfin A. Evaluating the correlation and performance of PROMIS to SRS questionnaires in adult and pediatric spinal deformity patients. Spine Deform 2019;7(1):118–24.

[37] Raad M, Harris AB, Puvanesarajah V, El-Dafrawy MH, Kebaish FN, Neuman BJ, et al. Preoperative patient expectations and pain improvement after adult spinal deformity surgery. J Neurosurg Spine SPI 2020;33(4):496–501. https://doi.org/10.3171/2020.3.SPINE191311.

[38] Lafage R, Ang B, Schwab F. Depression symptoms are associated with poor functional status among operative spinal deformity patients. Spine (Phila Pa 1976) 2021;46(7):447–56.