Andrew K. Chan, MD[*]
Thomas A. Wozny, MD[*]
Erica F. Bisson, MD, MPH[‡]
Brenton H. Pennicooke, MD, MS[*]
Mohamad Bydon, MD[§]
Steven D. Glassman, MD[¶]
Kevin T. Foley, MD[||]
Christopher I. Shaffrey, MD[#] [**]
Eric A. Potts, MD[‡‡]
Mark E. Shaffrey, MD[§§]
Domagoj Coric, MD[¶¶]
John J. Knightly, MD[||||]
Paul Park, MD[##]
Michael Y. Wang, MD[***]
Kai-Ming Fu, MD, PhD[‡‡‡]
Jonathan R. Slotkin, MD[§§§]
Anthony L. Asher, MD[¶¶]
Michael S. Virk, MD, PhD[‡‡‡]
Panagiotis Kerezoudis, MD, MS[§]
Mohammed A. Alvi, MBBS[§]
Jian Guan, MD[‡]
Regis W. Haid, MD[¶¶¶]
Praveen V. Mummaneni, MD, MBA[*]

[*]Department of Neurological Surgery, University of California, San Francisco, San Francisco, California, USA; [‡]Department of Neurosurgery, University of Utah, Salt Lake City, Utah, USA;
(Continued on next page)

Correspondence:
Andrew Kai-Hong Chan, MD,
Department of Neurological Surgery,
University of California, San Francisco,
505 Parnassus Ave, M779,
San Francisco, CA 94143, USA.
Email: andrewchanmd@gmail.com

Received, March 5, 2021.
Accepted, July 16, 2021.
Published Online, October 11, 2021.

Classifying Patients Operated for Spondylolisthesis: A K-Means Clustering Analysis of Clinical Presentation Phenotypes

BACKGROUND: Trials of lumbar spondylolisthesis are difficult to compare because of the heterogeneity in the populations studied.

OBJECTIVE: To define patterns of clinical presentation.

METHODS: This is a study of the prospective Quality Outcomes Database spondylolisthesis registry, including patients who underwent single-segment surgery for grade 1 degenerative lumbar spondylolisthesis. Twenty-four-month patient-reported outcomes (PROs) were collected. A k-means clustering analysis—an unsupervised machine learning algorithm—was used to identify clinical presentation phenotypes.

RESULTS: Overall, 608 patients were identified, of which 507 (83.4%) had 24-mo follow-up. Clustering revealed 2 distinct cohorts. Cluster 1 (high disease burden) was younger, had higher body mass index (BMI) and American Society of Anesthesiologist (ASA) grades, and globally worse baseline PROs. Cluster 2 (intermediate disease burden) was older and had lower BMI and ASA grades, and intermediate baseline PROs. Baseline radiographic parameters were similar ($P > .05$). Both clusters improved clinically ($P < .001$ all 24-mo PROs). In multivariable adjusted analyses, mean 24-mo Oswestry Disability Index (ODI), Numeric Rating Scale Back Pain (NRS-BP), Numeric Rating Scale Leg Pain, and EuroQol-5D (EQ-5D) were markedly worse for the high-disease-burden cluster (adjusted-$P < .001$). However, the high-disease-burden cluster demonstrated greater 24-mo improvements for ODI, NRS-BP, and EQ-5D (adjusted-$P < .05$) and a higher proportion reaching ODI minimal clinically important difference (MCID) (adjusted-$P = .001$). High-disease-burden cluster had lower satisfaction (adjusted-$P = .02$).

CONCLUSION: We define 2 distinct phenotypes—those with high vs intermediate disease burden—operated for lumbar spondylolisthesis. Those with high disease burden were less satisfied, had a lower quality of life, and more disability, more back pain, and more leg pain than those with intermediate disease burden, but had greater magnitudes of improvement in disability, back pain, quality of life, and more often reached ODI MCID.

KEY WORDS: Lumbar, Spondylolisthesis, Classification, Presentation, Clinical phenotypes, Quality Outcomes Database, Patient-reported outcomes

Neurosurgery 89:1033–1041, 2021 https://doi.org/10.1093/neuros/nyab355 www.neurosurgery-online.com

Degenerative lumbar spondylolisthesis affects an estimated 13.6% of the United States population.[1] For well-selected, symptomatic patients, surgery has been associated with clinical benefit.[2] Surgical approaches include decompression alone or decompression with fusion.[3-7] Multiple investigations[8-11]—including 3 randomized controlled trials[9,11,12]—have attempted to clarify the effectiveness of fusion for the treatment of lumbar spondylolisthesis but have come to inconsistent conclusions. Though there are multiple reasons for the different conclusions,[13] a critical issue lies in the different populations studied.

ABBREVIATIONS: MCID, minimal clinically important difference; **NASS,** North American Spine Society; **NRS-BP,** Numeric Rating Scale Back Pain; **NRS-LP,** Numeric Rating Scale Leg Pain; **ODI,** Oswestry Disability Index; **PRO,** patient-reported outcome; **QOD,** Quality Outcomes Database

The clinical heterogeneity of spondylolisthesis populations—despite radiographic attempts at classification such as by Meyerding classification[14]—is important to resolve in order to make appropriate comparisons across studies. Similar efforts have been undertaken for other spinal disorders, including cervical myelopathy.[15,16] A nuanced phenotypic understanding will permit spine surgeons, patients, and payers to draw appropriate conclusions from these investigations.

In order to define subgroups of patients with spondylolisthesis, it is important to investigate audited, multicenter datasets with large sample sizes. Although these large datasets have the potential to provide a nuanced differentiation of patient populations, making sense of such multidimensional information can be cumbersome. Clustering algorithms capable of automatically segmenting data are invaluable in revealing subgroups otherwise obscured by complexity. The k-means clustering, first proposed by MacQueen,[17] is one such machine learning algorithm, which optimally groups multidimensional clouds of data points using their relative spatial distribution.[18] This powerful technique has been successfully applied to a wide range of medical data including radiographic and clinical diagnostics in spine.[19-21]

We leverage the prospective Quality Outcomes Database (QOD) registry and focused spondylolisthesis dataset in order to identify distinct patterns of clinical presentation for patients operated for degenerative lumbar spondylolisthesis. This will aid future investigations in defining appropriate study populations and subgroups and may help clarify the utility of various surgical procedures as applied to unique spondylolisthesis phenotypes.

METHODS

Data Source

The QOD is a prospective, multicenter, multidisciplinary registry that contains data to measure the safety and quality of spinal surgery. Twelve of the highest-enrolling QOD sites were invited to participate in the Lumbar Spondylolisthesis Module/Study Group[8,11,22-32] This group was formed to investigate the effectiveness of fusion for grade 1 degenerative lumbar spondylolisthesis.[8,11] From July 2014 to June 2016, patients who underwent single-segment surgery for a diagnosis of Meyerding[14] grade 1 degenerative lumbar spondylolisthesis were enrolled. Surgeons at each participating site confirmed the diagnosis of grade 1 spondylolisthesis via an evaluation of standing or dynamic preoperative plain radiographs.

QOD exclusion criteria have been published previously.[33] All variables were audited for accuracy. Informed consent and institutional review board approval were obtained (University of California, San Francisco, IRB 16-20085).

Clinical Outcomes

We compared outcomes for identified clusters at 24 mo using validated questionnaires[34-37] (Table 1). An Oswestry Disability Index (ODI) improvement of 14.3 was considered the minimal clinically important difference (MCID).[22] Radiographic fusion and reduction in listhesis were determined by a neuroradiologist not affiliated with the study team.

Statistical Analysis

A k-means clustering algorithm—an unsupervised machine learning algorithm—was used to identify unique clusters of patients (Table 2). For univariate analyses, paired and unpaired *t*-tests and chi-square analyses were utilized with Yates' correction applied or Fisher exact test used where appropriate. Multivariable linear, multivariable logistic, and ordinal logistic regression models were fitted for adjusted analyses as appropriate. Factors reaching $P < .20$ on univariate comparisons of the cohorts were included as covariates in the multivariable models. This analysis was conducted using R 2.15.2 (R Foundation for Statistical Computing, Vienna, Austria). Missing values were imputed using the "missForest" R package. P-values were 2-tailed and an alpha of 0.05 was considered statistically significant.

RESULTS

Overall, 608 patients received single-segment surgery for Meyerding[14] grade 1 degenerative lumbar spondylolisthesis. A total of 507 patients (83.4%) had complete 24-mo follow-up and were included for further clustering analyses. The optimal number of clusters was found to be $k = 2$ when applying the silhouette technique. The robustness of this clustering result was further validated as the silhouette technique also found that the use of 2 clusters optimally partitioned the data for all principal component subspaces. A radar plot of the two patient cohorts resulting from k-means clustering of baseline patient parameters is shown in Figure. A total of 2 clusters were found to be optimal yielding 2 distinct cohorts of patients (Table 3): cluster 1 (high disease burden) represents a subset of patients with higher disease severity and cluster 2 (intermediate disease burden) represents a subset of patients with moderate disease severity.

Cluster 1 was defined by younger patients ($P < .001$), higher body mass index (BMI) ($P < .001$), higher American Society of Anesthesiologist (ASA) grades ($P < .001$), with a higher proportion of smokers ($P < .001$) and patients with anxiety ($P < .001$) and depression ($P < .001$). Clinically, cluster 1 had a higher proportion of patients presenting with back pain as a symptom ($P < .001$), a motor deficit ($P = .01$), who were not independently ambulatory ($P < .001$), and who had symptom durations >3 mo ($P = .04$). On validated assays of disability, pain, and quality of life, cluster 1 had globally worse ODI ($P < .001$), Numerical Rating Scale Back Pain (NRS-BP)

(Continued from previous page)

§Department of Neurologic Surgery, Mayo Clinic, Rochester, Minnesota, USA; ¶Norton Leatherman Spine Center, Louisville, Kentucky, USA; ||Department of Neurosurgery, Semmes-Murphey Neurologic and Spine Institute, University of Tennessee, Memphis, Tennessee, USA; #Department of Neurosurgery, Duke University, Durham, North Carolina, USA; **Department of Orthopedic Surgery, Duke University, Durham, North Carolina, USA; ‡‡Department of Neurological Surgery, Goodman Campbell Brain and Spine, Indianapolis, Indiana, USA; §§Department of Neurosurgery, University of Virginia, Charlottesville, Virginia, USA; ¶¶Neuroscience Institute, Carolina Neurosurgery & Spine Associates, Carolinas Healthcare System, Charlotte, North Carolina, USA; ||||Atlantic Neurosurgical Specialists, Morristown, New Jersey, USA; ##Department of Neurosurgery, University of Michigan, Ann Arbor, Michigan, USA; ***Department of Neurological Surgery, University of Miami, Miami, Florida, USA; ‡‡‡Department of Neurological Surgery, Weill Cornell Medical Center, New York, New York, USA; §§§Geisinger Health, Danville, Pennsylvania, USA; ¶¶¶Atlanta Brain and Spine Care, Atlanta, Georgia, USA

TABLE 1. PRO Measures

Oswestry Disability Index (ODI)	Range, 0 to 100, with higher scores indicating more disability related to back pain[34]
Numeric Rating Scale Back Pain (NRS-BP)	Range, 0 to 10, with higher scores indicating more back pain[35]
Numeric Rating Scale Leg Pain (NRS-LP)	Range, 0 to 10, with higher scores indicating more leg pain[35]
EuroQol-5D (EQ-5D)	Range, -0.11 to 1.0, with scores relaying a preference-based quality of life outcome metric and higher scores indicating less disability[36]
North American Spine Society (NASS) Satisfaction Questionnaire	Range, 1 to 4, with lower scores indicating more satisfaction with surgery[37]

TABLE 2. Methodology Employed for K-Means Clustering Analysis

The parameters used for clustering included age, BMI, ASA grade, baseline NRS-BP, baseline NRS-LP, the individual 5 component scores of the EQ-5D, and the individual component scores of the ODI except item #8 (9 components). ODI item #8 queries a patient's disability regarding pain affecting sex life and was not answered by a number of patients.
All parameters were z-score normalized across patients to account for differences in variance and scale across variable types and to ensure each parameter contributed equally to the clustering analysis.
The optimal number of clusters, *k*, was determined using the silhouette technique which quantifies the separation between clusters by measuring the distance between points in one cluster from those in neighboring clusters.
The optimal value of *k* was further validated by submitting the full dimensionality baseline parameter-space to a principal component analysis and applying this same silhouette technique to all principal component subspaces of iteratively increasing dimensionality as sorted by variance explained.
Once the optimal number of clusters was identified, a final clustering of the full dimensionality z-score normalized data was carried out to 1000 iterations to assure model convergence.
This clustering was repeated in its entirety a total of 100 times with different randomly generated centroid seeds to mitigate any initialization bias.

NRS-BP, Numeric Rating Scale Back Pain; NRS-LP, Numeric Rating Scale Leg Pain; EQ-5D, EuroQol-5D; ODI, Oswestry Disability Index.

($P < .001$), Numerical Rating Scale Leg Pain (NRS-LP) ($P < .001$), and EuroQol-5D[36] (EQ-5D) ($P < .001$). Surgically, there were higher proportion of cluster 1 patients who received fusions ($P = .003$) but otherwise similar proportions receiving minimally invasive surgeries and posterior vs alternative surgical approaches ($P > .05$ for both). Baseline radiographic parameters—mm of listhesis and proportion of dynamic listhesis—did not differ between clusters ($P > .05$).

Cluster 1 procedures were associated with higher blood loss ($P = .03$) and longer lengths of hospitalization ($P < .001$) (Table 4). A lower proportion of cluster 1 patients were discharged to home or home health care ($P = .003$).

Compared to baseline, both clusters improved at 24 mo for all outcomes ($P < .001$ for ODI, NRS-BP, NRS-LP, and EQ-5D). Table 5 reports the univariate comparisons of the 24-mo outcomes. In unadjusted analyses, cluster 1 had inferior mean 24-mo ODI ($P < .001$), NRS-BP ($P < .001$), NRS-LP ($P < .001$), and EQ-5D ($P < .001$). However, cluster 1 patients demonstrated greater 24-mo improvements (ie, change), for ODI ($P < .001$), NRS-BP ($P < .001$), and EQ-5D (<0.001). Cluster 1 patients also more often reached 24-mo ODI MCID ($P < .001$).

In multivariable analyses (Table 6)—to adjust for baseline differences between the clusters—we found that cluster 1 was associated with worse 24-mo ODI (adjusted-$P < .001$), NRS-BP (adjusted-$P < .001$), NRS-LP (adjusted-$P < .001$), EQ-5D (adjusted-$P < .001$), and North American Spine Society (NASS) satisfaction (adjusted-$P = .02$). However, cluster 1 was associated with superior 24-mo improvement for ODI change (adjusted-$P < .001$), NRS-BP change (adjusted-$P = .01$), and EQ-5D change (adjusted-$P < .001$). Furthermore, cluster 1 was associated with an increased odds of reaching ODI MCID (adjusted-$P = .001$).

DISCUSSION

In a machine learning k-means clustering analysis of 507 patients from the prospective QOD Spondylolisthesis Module dataset, we identified 2 distinct clinical presentations of patients who were operated for grade 1 degenerative lumbar spondylolisthesis. K-means clustering is a type of machine learning known as unsupervised learning in which characteristics within a dataset are identified and grouped based on spatial clustering of nearby data points into a multidimensional distribution. Cluster 1 (high disease burden) was defined by younger patients, higher BMI and ASA grades, and globally worse baseline patient-reported outcome (PRO) responses. Cluster 2 (intermediate disease burden) was defined by older patients, lower BMI and ASA grades, and globally intermediate baseline PRO responses. These separate phenotypes had different outcomes following surgery, with patients with high disease burden (cluster 1) having worse disability, back pain, leg pain, quality of life, and satisfaction at 24 mo. However, the high disease burden group experienced

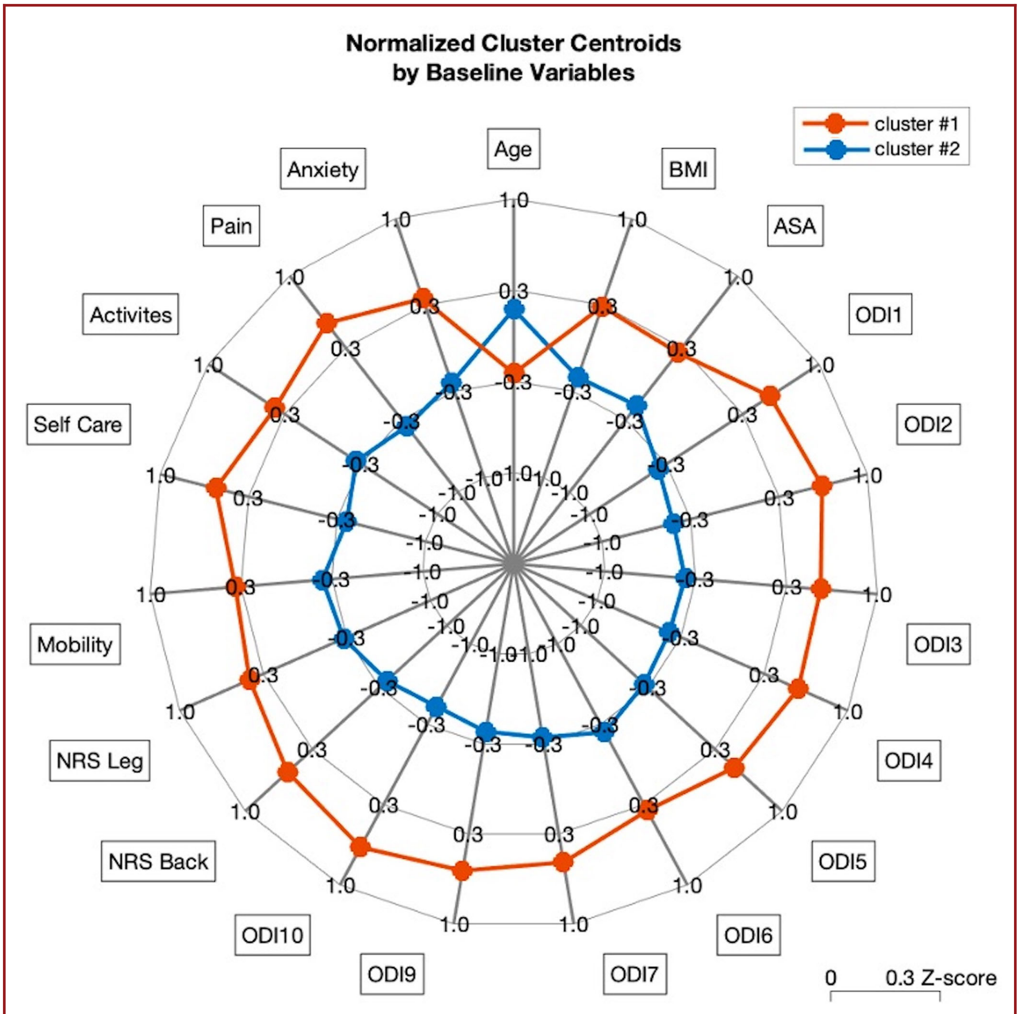

FIGURE. *The k-means clustering of baseline demographics and symptom scores. Radar plot of z-score normalized cluster centroids for k-means clustering (k = 2) of full dimensionality baseline data.*

greater improvements (ie, change score) with surgery for disability, back pain, and quality of life, and had a higher proportion of patients who were able to reach a MCID for disability. To facilitate future use of this classification scheme, a cluster calculator can be found here: [https://neurosurgicaltools.shinyapps.io/Spondylolisthesis_Phenotype_Calculator/].

Despite our cohort sharing the same radiographic classification of Meyerding[14] grade 1 lumbar spondylolisthesis, we identify the underlying clinical variability of these patients centered on 2 unique presentations—those with high and intermediate disease burden—each with notably different long-term outcomes. In this context, the inconsistent results of prior spondylolisthesis studies may be placed in appropriate context. In a recent systematic review of 37 studies investigating various management strategies for the treatment of lumbar spondylolisthesis,[38] only 1 of 6 probed topics (surgical vs nonoperative management) had sufficient evidence to warrant a grade A (ie, level 1 evidence with consistent findings) guideline recommendation. Otherwise, investigations involving the 5 remaining topics were inconsistent in findings or insufficient in quality to synthesize high-grade guideline recommendations and included studies on the utility of the addition of fusion to decompression, instrumented vs noninstrumented fusion, open vs minimally invasive surgery, the utility of interbody fusion in addition to posterolateral fusion alone, and the comparative effectiveness of various interbody approaches (ie, anterior vs lateral vs posterior/transforaminal). Considering the clinical variability observed in our study, it may not be surprising that many studies—grouping patients solely by radiographic classification—fail to draw similar conclusions. Future studies on these topics may be improved with specific adjustment

TABLE 3. Characteristics of Patients Undergoing Surgery for Grade 1 Lumbar Spondylolisthesis

	Cluster 1 high disease burden (n = 209)	Cluster 2 intermediate disease burden (n = 298)	P-value
Age (years), mean ± SD	59.2 ± 12.7	64.9 ± 11.2	<.001[a]
Female, n (%)	129 (61.7)	160 (53.7)	.07
BMI, mean ± SD	32.1 ± 6.8	28.6 ± 5.4	<.001[a]
Smoker, n (%)	35 (16.7)	18 (6.0)	<.001[a]
Diabetes mellitus, n (%)	42 (20.1)	43 (14.4)	.09
Coronary artery disease, n (%)	23 (11.0)	26 (8.7)	.39
Anxiety, n (%)	54 (25.8)	29 (9.7)	<.001[a]
Depression, n (%)	61 (29.2)	38 (12.8)	<.001[a]
Osteoporosis, n (%)	20 (9.6)	16 (5.4)	.07
mm of listhesis, mean ± SD	5.8 ± 3.0	6.4 ± 3.6	.21
Dynamic listhesis, n (%)	20 of 71 (28.2)	38 of 102 (37.3)	.21
Dominant presenting symptom, n (%)			<.001[a]
Back pain dominant	90 (43.1)	104 (34.9)	
Leg pain dominant	31 (14.8)	88 (29.5)	
Back pain = leg pain	88 (42.1)	106 (35.6)	
Motor deficit at baseline, n (%)	66 (31.6)	65 (21.8)	.01[a]
Independently ambulatory, n (%)	169 (80.9)	281 (94.3)	<.001[a]
Symptom duration, n (%)			.04[a]
<3 mo	4 (1.9)	10 (3.4)	
>3 mo	202 (96.7)	275 (92.3)	
ASA grade, n (%)			<.001[a]
1	4 (1.9)	19 (6.4)	
2	103 (49.3)	197 (66.1)	
3	97 (46.4)	82 (27.5)	
4	5 (2.4)	0 (0)	
Hispanic or Latino, n (%)	19 (9.1)	9 (3.0)	.003[a]
4 yr of college education or more, n (%)	59 (28.2)	155 (52.0)	<.001[a]
Employed or employed and on leave, n (%)	90 (43.2)	142 (47.6)	.34
Private insurance, n (%)	114 (54.5)	155 (52.0)	.57
Use of worker's compensation, n (%)	6 (2.9)	10 (3.4)	.76
ODI, baseline, mean ± SD	60.5 ± 12.2	35.6 ± 12.3	<.001[a]
NRS back pain, baseline, mean ± SD	8.3 ± 1.8	5.5 ± 2.8	<.001[a]
NRS leg pain, baseline, mean ± SD	7.8 ± 2.3	5.7 ± 2.9	<.001[a]
EQ-5D, baseline, mean ± SD	0.38 ± 0.19	0.66 ± 0.16	<.001[a]
>2-yr follow-up rate, n (%)	168 (80.4)	266 (89.3)	.005[a]
Surgical characteristics			
Decompression alone, n (%)	38 (18.2)	89 (29.9)	.003[a]
Decompression and fusion, n (%)	171 (81.8)	209 (70.1)	.003[a]
Use of minimally invasive techniques, n (%)	92 (44.0)	136 (45.6)	.72
Surgical approach			.51
Posterior only, n (%)	198 (94.7)	275 (92.3)	
Anterior only, n (%)	3 (1.4)	10 (3.4)	
Lateral only, n (%)	2 (1.0)	5 (1.7)	
Two-stage, n (%)	6 (2.9)	8 (2.7)	

BMI, body mass index; ASA, American Society of Anesthesiologists; MI, minimally invasive surgery; ODI, Oswestry Disability Index; NRS, Numeric Rating Scale; EQ-5D, EuroQol-5D; SD, standard deviation. Values do not add up to 100% where there is missing data.
[a]Denotes a significant difference with *P*-value < .05.

for these clinical presentation clusters. Patients in each cluster may be analyzed separately or cluster identity may be factored into propensity-matched or multivariable regression models.

As opposed to disease classification schemes that rely solely on a few clinical characteristics, we use a sophisticated machine learning method, k-means clustering, to identify unique phenotypes of patients operated for lumbar spondylolisthesis. Applying this analysis to a large registry dataset permits the identification of patterns that may not be otherwise obvious. Such machine learning approaches may uncover unique patient profiles and—

TABLE 4. Operative Data for Patients Undergoing Surgery for Grade 1 Lumbar Spondylolisthesis

	Cluster 1 high disease burden (n = 209)	Cluster 2 intermediate disease burden (n = 298)	*P*-value
Estimated blood loss (ml), mean ± SD	188.5 ± 197.5	152.3 ± 166.5	.03[a]
Operative time (minutes), mean ± SD	169.9 ± 70.3	176.5 ± 95.3	.40
Length of hospitalization (days), mean ± SD	3.2 ± 1.9	2.3 ± 1.6	<.001[a]
Discharge disposition			
Home or home health care, n (%)	184 (88.0)	284 (95.3)	.003[a]
Readmissions, 3 mo, n (%)	4 (1.9)	3 (1.0)	.63
Reoperations, 24-36 mo follow-up, n (%)	11 (5.3)	17 (5.7)	.83
Complications, 30 d, n (%)	16 (7.7)	16 (5.4)	.30

SD, standard deviation; ml, milliliters; n/a, not applicable. [a]Denotes a significant difference with *P*-value < .05.

TABLE 5. Univariate Comparison of Outcomes for Patients Undergoing Surgery for Grade 1 Lumbar Spondylolisthesis

	Cluster 1 high disease burden (n = 209)	Cluster 2 intermediate disease burden (n = 298)	*P*-value
ODI			
MCID	n = 168	n = 266	<.001[a]
Reached	132 (78.6)	162 (60.9)	
Not reached	36 (21.4)	104 (39.1)	
ODI, 24 mo, mean ± SD	29.7 ± 21.1	16.0 ± 15.3	<.001[a]
ODI, change, mean ± SD	−30.9 ± 21.5	−19.7 ± 17.9	<.001[a]
NRS back pain			
NRS back pain, 24 mo, mean ± SD	4.1 ± 3.1	2.6 ± 2.7	<.001[a]
NRS back pain, change, mean ± SD	−4.0 ± 3.5	−2.9 ± 3.3	<.001[a]
NRS leg pain			
NRS leg pain, 24 mo, mean ± SD	3.5 ± 3.4	1.8 ± 2.7	<.001[a]
NRS leg pain, change, mean ± SD	−4.2 ± 4.0	−3.9 ± 3.6	.40
EQ-5D			
EQ-5D, 24 mo, mean ± SD	0.69 ± 0.21	0.82 ± 0.17	<.001[a]
EQ-5D, change, mean ± SD	+0.31 ± 0.25	+0.16 ± 0.22	<.001[a]
NASS satisfaction, n (%)	n = 167	n = 258	.054
1	92 (55.1)	174 (67.4)	
2	42 (25.1)	46 (17.8)	
3	16 (9.6)	14 (5.4)	
4	17 (10.2)	24 (9.3)	

ODI, Oswestry Disability Index; MCID, minimal clinically important difference; NRS, Numeric Rating Scale; EQ-5D, EuroQol-5D; NASS, North American Spine Society; SD, standard deviation. [a]Denotes a significant difference with *P*-value < .05.

aside from helping to standardize research investigations—may permit the tailoring of care to the individual patient.[39,40] This avoids the somewhat crude reliance on published literature, drawn from heterogeneous spondylolisthesis populations, that may not be entirely relevant to a unique patient. Here, we provide unique spondylolisthesis patient phenotypes that may be helpful for counseling, including informed consent and expectation setting with regard to clinical outcomes. These clusters—high vs intermediate disease burden—are readily replicated and identified by the provision of 19 features to a calculator of cluster centroid: age, BMI, ASA grade, 9 of 10 component scores of the baseline ODI, 5 component scores of the baseline EQ-5D, baseline NRS-BP score, and baseline NRS-LP score. These features may be collected in an automated or nearly automated fashion as patients complete baseline PRO questionnaires at home or in the clinic. Workflows may be established—including the use of the electronic medical record—to readily provide this information to the surgeon or clinical research investigator. Furthermore, this information can be readily entered into the publicly available online calculator.

Our 2 identified clusters may represent 2 distinct patients that spine surgeons may recognize from clinic. The first may represent the younger patient with a higher back pain and foraminal

TABLE 6. Impact of High Disease Burden on 24-Month Outcomes Following Surgery for Grade 1 Lumbar Spondylolisthesis, Compared to Intermediate Disease Burden

	Adjusted[a] β coefficients (95% CI)	P-value
ODI, 24 mo	12.7 (9.1 to 16.3)	<.001[b]
ODI change, 24 mo	− 9.2 (−13.0 to 5.5)	<.001[b]
NRS back pain, 24 mo	1.3 (0.7 to 1.8)	<.001[b]
NRS back pain change, 24 mo	− 0.8 (−1.4 to −0.2)	.01[b]
NRS leg pain, 24 mo	1.5 (1.0 to 2.1)	<.001[b]
NRS leg pain change, 24 mo	− 0.4 (−1.1 to 0.3)	.30
EQ-5D, 24 mo	− 0.10 (−0.14 to −0.06)	<.001[b]
EQ-5D change, 24 mo	+0.16 (+0.12 to +0.21)	<.001[b]
	Adjusted[a] odds ratio (95% CI)	
ODI MCID, 24 mo	2.1 (1.3 to 3.5)	.001[b]
NASS satisfaction, 24 mo	0.6 (0.4 to 0.9)	.02[b]

ODI, Oswestry Disability Index; NRS, Numerical Rating Scale; EQ-5D, EuroQol-5D; BMI, body mass index; MCID, minimal clinically important difference; NASS, North American Spine Society.

β Coefficients for change scores are reported such that a negative value for ODI, NRS back pain, and NRS leg pain, and a positive value for EQ-5D represents superior outcomes at 24 mo.

Odds ratios (OR) are reported such that an OR > 1.0 for NASS satisfaction represents greater satisfaction at 24 mo.

[a]Multivariable models adjusted for factors with $P < .20$ on univariate comparisons of clusters 1 and 2.

[b]Denotes statistical significance with a P-value < .05.

stenosis dominant presentation—a presentation with disease burden more clearly reflected in our generic spine PRO measures. The second may represent the older patient with a primary complaint of lumbar stenosis and neurogenic claudication—symptoms that are less well reflected in baseline ODI, back pain, and leg pain. Patients with neurogenic claudication do quite well with surgery,[41-43] which is reflected in cluster 2 having significantly higher NASS satisfaction scores than cluster 1. As large, multicenter datasets accumulate, these clusters may be further refined and validated.

Although anxiety was included as a variable in clustering, it is interesting to note the significant differences in anxiety and depression between clusters with increased anxiety and depression seen in the high disease burden group. Proof of directional causal relationships is beyond the scope of this dataset, but we note the possible contribution of a biopsychological feedback loop in which disability secondary to spinal pathology contributes to negative psychological effects, which in turn exacerbates the cognitive aspects of pain perception and further worsens disability.[44-46] As a recent study highlights, there is a deleterious association between depression and anxiety on short term outcomes following surgery for grade 1 degenerative lumbar spondylolisthesis.[23] Further research into comorbid psychological conditions and their targeted treatment may be conducted to help to improve the treatment of spinal pathologies.

Limitations

This study has several limitations. First, this is a retrospective analysis of prospective registry data and should be interpreted accordingly. These limitations include, but are not limited to, issues of loss to follow-up.[47] The QOD dataset analyzed here had a 2-yr follow-up rate of 85% and factors associated with loss of follow-up have been explored in a recent study.[48] Ultimately, 83% of patients had complete data for all analyzed variables and could thus be subjected to cluster analysis. Although higher follow-up rates are always preferable, our rate remains an acceptable one in the context of prospective registry, 2-yr follow-up (>80%).[49] Second, this analysis defines clinical presentations of patients with Meyerding[14] grade 1 degenerative lumbar spondylolisthesis who ultimately received surgery. Therefore, the findings may not apply to patients with higher grades of spondylolisthesis, nondegenerative etiologies, or not receiving surgery (ie, who may be considered to have low disease burden). Third, this study was designed to compare outcomes of patients based on cluster identity and was not powered to draw strong conclusions regarding the multitiered interaction of cluster identity, specific type of surgical intervention (eg, decompression only vs decompression and fusion, minimally invasive vs open, etc), and outcome. This further subdivision of groupings, compounded by the heterogeneity of specific surgical intervention undertaken, and unequal distribution of patients across groups resulted in some groups being undersized for reliable statistical comparison. Fourth, floor and ceiling effects may contribute to our outcome findings. However, because we report widely used, validated PROs, our findings should parallel those encountered by spine surgeons using these metrics. Although our cluster analysis was undertaken on a large, multicenter registry, the utility of our findings remains an open question. One goal of our investigation is to provide more homogeneous sets of patients to study, thereby increasing the external validity of findings in future studies on populations with spondylolisthesis. However, we note that these clusters are not intended to replace other important methodologies necessary for high-quality investigation (eg, randomization). Further validation of our defined clusters in other spondylolisthesis populations—such as in the Spinal Laminectomy Versus Instrumented Pedicle Screw Fusion (SLIP II) Registry (clinicaltrials.gov identifier: NCT03570801)—may be pursued.

CONCLUSION

Using a machine learning clustering analysis of the prospective QOD Spondylolisthesis Module dataset, we identified two distinct groups of patients—those with high vs intermediate disease burden—who ultimately present for single-segment surgery for lumbar spondylolisthesis. At 24 mo, those with high disease burden were less satisfied, had a lower quality of life, and more severe disability, back pain, and leg pain compared to those with intermediate disease burden. However, compared to those with intermediate disease burden, the high disease burden group experienced greater improvements (ie, change score) with surgery for disability, back pain, and quality of life and had a higher proportion ultimately able to reach an ODI MCID. Because these clusters are associated with significantly different outcomes, it may be helpful for future investigations to consider these clinical phenotypes as distinct entities. The use of machine learning algorithms to mine large, curated datasets may further help reveal previously unrecognized clinical endophenotypes across other types of neurosurgical, orthopedic, and spinal disorders. For spondylolisthesis in particular, we posit that further refinement and prospective validation of subgroup analyses with larger and richer datasets could ultimately be used by patients in helping to set expectations, by researchers for consideration in future studies, and by surgeons in preoperative planning and counseling.

Funding

The Neurosurgery Research and Education Foundation provided financial support for this work.

Disclosures

Dr Chan receives nonstudy-related research support from Orthofix Inc. Dr Bisson is a consultant for nView medical and MiRus, and also has direct stock ownership on those companies. She receives clinical or research support for the study described (includes equipment or material) from the Neurosurgery Research and Education Foundation (NREF). Dr Glassman is an employee of Norton Healthcare; is a consultant for K2M and Medtronic; is a patent holder with Medtronic, from which he receives royalties; and receives clinical or research support for the study described (includes equipment or material) from NuVasive. Dr Foley is a consultant for Medtronic; has direct stock ownership in Digital Surgery Systems, Discgenics, DuraStat, LaunchPad Medical, Medtronic, NuVasive, nView medical, Practical Navigation/Fusion Robotics, Spine Wave, TDi, and Triad Life Sciences; is a patent holder with Medtronic and NuVasive; and is a member of the board of directors of Digital Surgery Systems, Discgenics, DuraStat, LaunchPad Medical, nView medical, Practical Navigation/Fusion Robotics, TDi, and Triad Life Sciences. Dr Shaffrey has direct stock ownership with NuVasive; is a patent holder with NuVasive, Medtronic, Zimmer Biomet, and SI Bone; and is a consultant for NuVasive, Medtronic, and SI Bone. Dr Potts is a patent holder with Medtronic. Dr Coric is a consultant for and holds direct stock ownership in Spine Wave; is a consultant for Stryker; is a consultant for Medtronic; is a consultant for and holds direct stock ownership in Premia Spine; and holds direct stock ownership in Spinal Kinetics. Dr Knightly is on the NPA board of directors. Dr Park is a consultant for Globus Medical and NuVasive; receives royalties from Globus Medical; and receives support of a nonstudy-related clinical or research effort that he oversees from Pfizer and Vertex. Dr Wang is a consultant for DePuy Synthes Spine, K2M, Stryker, and Spineology; has direct stock ownership in ISD and Medical Devices Partners; and is a patent holder with DePuy Synthes Spine. Dr Fu is a consultant for SI Bone, Globus Medical, and Johnson & Johnson. Dr Slotkin is a consultant for Medtronic and Stryker. Dr Virk is a consultant for and received honorarium from DePuy Synthes Spine Inc, BrainLab Inc, and Globus Medical. Dr Haid has direct stock ownership in Globus Medical, NuVasive, Paradigm Spine, Spine Universe (Vertical Health), and Spine Wave. He also receives royalties for IP; Globus Lateral and TLIF Interbody Implants; Medtronic Atlantis; Venture Anterior Plates; Medtronic Prestige ST and LP; NuVasive ALIF; Post Pedicle Screw Reline; and multiple textbooks. He sits on the board of directors for the AANS, Lumbar Spine Research Society, and NREF as well. Dr Mummaneni is a consultant for DePuy Synthes, Globus Medical, and Stryker; has direct stock ownership in Spinicity-ISD; receives clinical or research support for the study described (includes equipment or material) from NREF; receives support of a nonstudy-related clinical or research effort that he oversees from ISSG and AOSpine; and receives royalties from Thieme Publishing, Springer Publishing, and DePuy Synthes.

REFERENCES

1. Kalichman L, Kim DH, Li L, Guermazi A, Berkin V, Hunter DJ. Spondylolysis and spondylolisthesis: prevalence and association with low back pain in the adult community-based population. *Spine (Phila Pa 1976)*. 2009;34(2):199-205.
2. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med*. 2007;356(22):2257-2270.
3. Dhall SS, Wang MY, Mummaneni PV. Clinical and radiographic comparison of mini-open transforaminal lumbar interbody fusion with open transforaminal lumbar interbody fusion in 42 patients with long-term follow-up. *J Neurosurg Spine*. 2008;9(6):560-565.
4. Wang JC, Mummaneni PV, Haid RW. Current treatment strategies for the painful lumbar motion segment: posterolateral fusion versus interbody fusion. *Spine (Phila Pa 1976)*. 2005;30(16 Suppl):S33-S43.
5. Rosenberg WS, Mummaneni PV. Transforaminal lumbar interbody fusion: technique, complications, and early results. *Neurosurgery*. 2001;48(3):569-575; discussion 574-565.
6. Ghogawala Z, Shaffrey CI, Asher AL, et al. The efficacy of lumbar discectomy and single-level fusion for spondylolisthesis: results from the NeuroPoint-SD registry: clinical article. *J Neurosurg Spine*. 2013;19(5):555-563.
7. Cheng JS, Park P, Le H, Reisner L, Chou D, Mummaneni PV. Short-term and long-term outcomes of minimally invasive and open transforaminal lumbar interbody fusions: is there a difference? *Neurosurg Focus*. 2013;35(2):E6.
8. Chan AK, Bisson EF, Bydon M, et al. Laminectomy alone versus fusion for grade 1 lumbar spondylolisthesis in 426 patients from the prospective Quality Outcomes Database. *J Neurosurg Spine*. 2018;30(2):234-241.
9. Forsth P, Olafsson G, Carlsson T, et al. A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. *N Engl J Med*. 2016;374(15):1413-1423.
10. Ghogawala Z, Dziura J, Butler WE, et al. Laminectomy plus fusion versus laminectomy alone for lumbar spondylolisthesis. *N Engl J Med*. 2016;374(15):1424-1434.
11. Bisson EF, Guan J, Bydon M, et al. Patient-reported outcome improvements at 24-month follow-up after fusion added to decompression for grade 1 degenerative lumbar spondylolisthesis: a multicenter study using the Quality Outcomes Database. *J Neurosurg Spine*. 2021;35(1):42-51.
12. Austevoll IM, Hermansen E, Fagerland MW, et al. Decompression with or without fusion in degenerative lumbar spondylolisthesis. *N Engl J Med*. 2021;385(6):526-538.
13. Ghogawala Z, Resnick DK, Glassman SD, Dziura J, Shaffrey CI, Mummaneni PV. Randomized controlled trials for degenerative lumbar spondylolisthesis: which patients benefit from lumbar fusion? *J Neurosurg Spine*. 2017;26(2):260-266.

14. Meyerding HW. Diagnosis and roentgenologic evidence of spondylolisthesis. *Radiology*. 1933;20(2):108-120.
15. Benzel EC, Lancon J, Kesterson L, Hadden T. Cervical laminectomy and dentate ligament section for cervical spondylotic myelopathy. *J Spinal Disord*. 1991;4(3):286-295.
16. Nurick S. The natural history and the results of surgical treatment of the spinal cord disorder associated with cervical spondylosis. *Brain*. 1972;95(1):101-108.
17. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press; 1967.
18. Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognit Lett*. 2010;31(8):651-666.
19. Larhmam MA, Benjelloun M, Mahmoudi S. Vertebra identification using template matching modelmp and K-means clustering. *Int J Comput Assist Radiol Surg*. 2014;9(2):177-187.
20. Sabaghian S, Dehghani H, Batouli SAH, Khatibi A, Oghabian MA. Fully automatic 3D segmentation of the thoracolumbar spinal cord and the vertebral canal from T2-weighted MRI using K-means clustering algorithm. *Spinal Cord*. 2020;58(7):811-820.
21. Melo Riveros NA, Cardenas Espitia BA, Aparicio Pico LE. Comparison between K-means and self-organizing maps algorithms used for diagnosis spinal column patients. *Inform Med Unlocked*. 2019;16:100206.
22. Asher AL, Kerezoudis P, Mummaneni PV, et al. Defining the minimum clinically important difference for grade I degenerative lumbar spondylolisthesis: insights from the Quality Outcomes Database. *Neurosurg Focus*. 2018;44(1):E2.
23. Kashlan O, Swong K, Alvi MA, et al. Patients with a depressive and/or anxiety disorder can achieve optimum long term outcomes after surgery for grade 1 spondylolisthesis: analysis from the Quality Outcomes Database (QOD). *Clin Neurol Neurosurg*. 2020;197:106098.
24. Chan AK, Bisson EF, Bydon M, et al. Women fare best following surgery for degenerative lumbar spondylolisthesis: a comparison of the most and least satisfied patients utilizing data from the Quality Outcomes Database. *Neurosurg Focus*. 2018;44(1):E3.
25. Chan AK, Bisson EF, Bydon M, et al. A comparison of minimally invasive transforaminal lumbar interbody fusion and decompression alone for degenerative lumbar spondylolisthesis. *Neurosurg Focus*. 2019;46(5):E13.
26. Chan AK, Bisson EF, Bydon M, et al. Obese patients benefit, but do not fare as well as nonobese patients, following lumbar spondylolisthesis surgery: an analysis of the Quality Outcomes Database. *Neurosurgery*. 2020;86(1):80-87.
27. Chan AK, Bisson EF, Fu KM, et al. Sexual dysfunction: prevalence and prognosis in patients operated for degenerative lumbar spondylolisthesis. *Neurosurgery*. 2019;87(2):200-210.
28. Karsy M, Chan AK, Mummaneni PV, et al. Outcomes and complications with age in spondylolisthesis: an evaluation of the elderly from the Quality Outcomes Database. *Spine (Phila Pa 1976)*. 2020;45(14):1000-1008.
29. Mummaneni PV, Bisson EF, Kerezoudis P, et al. Minimally invasive versus open fusion for grade I degenerative lumbar spondylolisthesis: analysis of the Quality Outcomes Database. *Neurosurg Focus*. 2017;43(2):E11.
30. Mummaneni PV, Bydon M, Alvi MA, et al. Predictive model for long-term patient satisfaction after surgery for grade I degenerative lumbar spondylolisthesis: insights from the Quality Outcomes Database. *Neurosurg Focus*. 2019;46(5):E12.
31. Wang MY, Tran S, Brusko GD, et al. Less invasive spinal deformity surgery: the impact of the learning curve at tertiary spine care centers [published online ahead of print: August 23, 2019]. *J Neurosurg Spine*. doi:10.3171/2019.6.SPINE19531.
32. Sherrod BA, Mummaneni PV, Alvi MA, et al. Regional variance in disability and quality-of-life outcomes after surgery for grade I degenerative lumbar spondylolisthesis: a Quality Outcomes Database analysis. *World Neurosurg*. 2020;138:e336-e344.
33. McGirt MJ, Speroff T, Dittus RS, Harrell FE Jr, Asher AL. The National Neurosurgery Quality and Outcomes Database (N2QOD): general overview and pilot-year project description. *Neurosurg Focus*. 2013;34(1):E6.
34. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy*. 1980;66(8):271-273.
35. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? *Pain*. 1994;58(3):387-392.
36. EuroQol G. EuroQol–a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199-208.
37. Daltroy LH, Cats-Baril WL, Katz JN, Fossel AH, Liang MH. The North American spine society lumbar spine outcome assessment instrument: reliability and validity tests. *Spine (Phila Pa 1976)*. 1996;21(6):741-748.
38. Chan AK, Sharma V, Robinson LC, Mummaneni PV. Summary of guidelines for the treatment of lumbar spondylolisthesis. *Neurosurg Clin N Am*. 2019;30(3):353-364.
39. Amara D, Mummaneni PV, Ames CP, et al. Treatment of only the fractional curve for radiculopathy in adult scoliosis: comparison to lower thoracic and upper thoracic fusions. *J Neurosurg Spine*. 2019;30(4):506-514.
40. Joshi RS, Lau D, Ames CP. Artificial intelligence and the future of spine surgery introduction. *Neurospine*. 2019;16(4):637-639.
41. Amundsen T, Weber H, Nordal HJ, Magnaes B, Abdelnoor M, Lilleas F. Lumbar spinal stenosis: conservative or surgical management? A prospective 10-year study. *Spine (Phila Pa 1976)*. 2000;25(11):1424-1436; discussion 1435-1426.
42. Malmivaara A, Slatis P, Heliovaara M, et al. Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial. *Spine (Phila Pa 1976)*. 2007;32(1):1-8.
43. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the spine patient outcomes research trial. *Spine (Phila Pa 1976)*. 2010;35(14):1329-1338.
44. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. *Arch Intern Med*. 2003;163(20):2433-2445.
45. Boersma K, Linton SJ. Psychological processes underlying the development of a chronic pain problem: a prospective study of the relationship between profiles of psychological variables in the fear-avoidance model and disability. *Clin J Pain*. 2006;22(2):160-166.
46. Haythornthwaite JA, Sieber WJ, Kerns RD. Depression and the chronic pain experience. *Pain*. 1991;46(2):177-184.
47. Dettori JR. Loss to follow-up. *Evid Based Spine Care J*. 2011;2(01):7-10.
48. Bisson EF, Mummaneni PV, Knightly J, et al. Assessing the differences in characteristics of patients lost to follow-up at 2 years: results from the Quality Outcomes Database study on outcomes of surgery for grade I spondylolisthesis. *J Neurosurg Spine*. 2020;33(5):643-651.
49. Luckmann R. *Evidence-Based Medicine: How to Practice and Teach EBM*. Sackett David L. SES, Richardson W. Scott, Rosenberg William, Haynes R. Brian, ed. Vol 16. 2nd ed. Churchill Livingstone; 2000:155-156.

Acknowledgments

We would like to thank all of the site research coordinators for their help with data extraction and validation.