Spine

SPINE Volume 43, Number 5, pp 339–347

DEFORMITY

Operative Management of Adult Spinal Deformity Results in Significant Increases in QALYs Gained Compared to Nonoperative Management

Analysis of 479 Patients With Minimum 2-Year Follow-Up

Justin K. Scheer, BS,[*] Richard Hostin, MD,[†] Chessie Robinson, MA,[‡] Frank Schwab, MD,[§] Virginie Lafage, PhD,[‡] Douglas C. Burton, MD,[¶] Robert A. Hart, MD,[||] Michael P. Kelly, MD,[**] Malla Keefe, BS,[††] David Polly, MD,[‡‡] Shay Bess, MD,[§§] Christopher I. Shaffrey, MD,[¶¶] Justin S. Smith, MD, PhD,[¶¶] and Christopher P. Ames, MD[††], the International Spine Study Group

Study Design. Retrospective review of prospective multicenter adult spinal deformity (ASD) database.

Objective. To compare the quality-adjusted life years (QALYs) between operative and nonoperative treatments for ASD patients.

Summary of Background Data. Operative management of ASD repeatedly demonstrates improvements in HRQOL over nonoperative treatment. However, little is reported regarding QALY improvements after surgical correction of ASD.

Methods. Inclusion criteria: ≥18 years, ASD. Health utility values were calculated from SF6D scores and used to calculate QALYs at minimum 2 years from the baseline utility value as well as at 1, 2, and 3 years for the available patients. A 1:1 propensity score matching using six baseline variables was conducted to account for the nonrandom distribution of operative and nonoperative treatments.

Results. Four hundred seventy-nine patients were included (OP:258, 70.7%, NONOP:221, 47.1%). One hundred fifty-one (OP:90, NONOP:61) had complete 1, 2, and 3 year data available for QALY trending. Unmatched results are not listed in the abstract. Mean baseline utility scores were statistically similar between the matched groups (OP: 0.609 ± 0.093, NONOP: 0.600 ± 0.091, $P = 0.6401$) and at 2 year min postop mean OP QALY was greater than NONOP (1.377 ± 0.345 *vs.* 1.256 ± 0.286, respectively, $P < 0.01$). For the subanalysis cohort, mean OP QALYs at 1, 2, and 3 years postoperative were all significantly greater than NONOP, $P < 0.03$ for all (1 yr: 0.651 ± 0.089 *vs.* 0.61 ± 0.079, 2 yr: 1.29 ± 0.157 *vs.* 1.189 ± 0.141, and 3 yr: 1.903 ± 0.235 *vs.* 1.749 ± 0.198, respectively). Matched OP had a larger QALYs gained (from baseline) at 2 year minimum postoperative (0.112 ± 0.243 *vs.* 0.008 ± 0.195, $P < 0.01$). For subanalysis of patients with complete 1 to 3 years data, OP had a significantly larger QALYs gained at 1, 2, and 3 years compared with NONOP: 1 year (0.073 ± 0.121 *vs.* 0.029 ± 0.082, $P = 0.0447$), 2 years (0.167 ± 0.232 *vs.* 0.036 ± 0.173, $P = 0.0030$), and 3years (0.238 ± 0.379 *vs.* 0.059 ± 0.258, $P < 0.01$).

Conclusion. The operative treatment of adult spinal deformity results in significantly greater mean QALYs and QALYs gained at minimum 2 years postop as well as at the 1-, 2-, and 3-year time points compared with nonoperative management.

Key words: adult spinal deformity, ASD, nonoperative treatment, QALY, quality-adjusted life-year.

Level of Evidence: 3

Spine 2018;43:339–347

The operative management of adult spinal deformity (ASD) can exceed $100,000 with a reported mean cost of $72,034 (range $10,768–$187,458).[1] In a

From the [*]School of Medicine, University of California, San Diego, La Jolla, CA; [†]Department of Orthopaedic Surgery, Baylor Scoliosis Center, Plano, TX; [‡]Center for Clinical Effectiveness, Baylor Scott & White Health, Dallas, TX; [§]Department of Orthopaedic Surgery, NYU Hospital for Joint Diseases, New York, NY; [¶]Department of Orthopaedic Surgery, University of Kansas Medical Center, Kansas City, KS; [||]Department of Orthopaedic Surgery, Oregon Health and Science University, Portland, OR; [**]Department of Orthopedic Surgery, Washington University, St. Lois, MO; [††]Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA; [‡‡]Department of Orthopaedic Surgery, University of Minnesota, Minneapolis, MN; [§§]Rocky Mountain Hospital for Children, Denver, CO; and [¶¶]Department of Neurosurgery, University of Virginia Health System, Charlottesville, VA.

Acknowledgment date: December 3, 2015. First revision date: February 13, 2016. Acceptance date: April 4, 2016.

The manuscript submitted does not contain information about medical device(s)/drug(s).

The International Spine Study Group Foundation, through which this study was conducted, is funded through research grants from DePuy Spine and individual donations.

Relevant financial activities outside the submitted work: board membership, consultancy, employment, expert testimony, grants, patents, royalties, payment for lectures, stocks, travel/accommodations/meeting expenses.

Address correspondence and reprint requests to Justin K. Scheer, BS, School of Medicine, University of California, San Diego, La Jolla, CA 92093; E-mail: jscheer@ucsd.edu

DOI: 10.1097/BRS.0000000000001626

healthcare environment currently in flux with a greater emphasis on costs, this is of critical importance; especially since the prevalence of ASD in the general population has been reported to be up to 32%[2–5] and as high as 68% in the older population.[6] Given the high cost of operative treatment, a rapidly growing patient population, and significant healthcare reform, studies using quality-adjusted life years (QALYs) are critical to determining the appropriate care as well as aid in developing ways to reduce costs.

However, there are currently few data on the disease burden associated with ASD as measured by QALYs and those studies that do exist have only just recently been published.[1,7–11] Of those, two studies specifically attempted to quantify the cost-effectiveness of operative over nonoperative treatment for hypothetical future QALYs by projecting them out to 5 and 10 years from surgery.[8,10] Both studies determined that surgery for ASD was cost-effective at the final time point assessed; however, these conclusions were drawn from hypothetical QALYs.

To date, no study has directly compared the QALYs gained between operative and nonoperative treatments for adult spinal deformity patients. Thus, the purpose of the present study was to compare the QALYs gained between operative and nonoperative treatments for patients with adult spinal deformity at a minimum of 2 years postoperative and trend these QALYs yearly for 3 years postoperative in the available patients.

MATERIALS AND METHODS

Patient Population

This study is a retrospective review of a prospective multicenter ASD database, which is composed of 11 sites across the United States. All patients were enrolled from 2008 to 2012 and enrolled into an Institutional Review Board-approved protocol by each site. Inclusion criteria for the database were: age >18 years and presence of spinal deformity, as defined by scoliosis Cobb angle $\geq 20^\circ$, SVA ≥ 5 cm, PT $\geq 25^\circ$, and/or thoracic kyphosis (TK) $\geq 60^\circ$. Exclusion criteria included spinal deformity of a neuromuscular etiology and presence of active infection or malignancy.

Patients were divided into nonoperative (NONOP) and operative (OP) cohorts at the time of enrollment. The decision to pursue operative management was determined between each individual patient and surgeon. The majority of ASD patients seen in our clinics have received nonoperative treatment and they are being seen for possible surgical intervention. Patients presenting with a progressive neurologic compromise, myelopathy, or bowel/bladder incontinence were generally advised to pursue operative treatment. The remaining patients were counseled regarding both operative and nonoperative management options. Those that elected nonoperative care were not enrolled into a standardized nonoperative treatment protocol, but rather care was catered to the patients' goals. Patients in both operative and nonoperative treatment groups were followed at regular intervals. Nonoperative patients who elected for operative treatment after study enrollment were categorized as operative and analyzed as such.

Data Collection, Radiographic Assessment, and HRQOL

The demographic and clinical data included patient age, sex, body mass index (BMI), Charlson Comorbidity Index (CCI),[12] Anesthesiologists (ASA) physical status classification. Surgical data collected included: approach, levels fused, presence and type and number of interbody fusions, osteotomy types, operating room time (OR time), estimated blood loss (EBL), length of hospital stay (LOS), and major and minor complications as defined by Glassman *et al*.

Standardized HRQOL measures included the Oswestry Disability Index (ODI), Short-Form-36 (SF-36), and Scoliosis Research Society-22 (SRS-22). A numeric rating scale (NRS) score ranging from 0 (no pain) to 10 (most unbearable pain) was collected for back and leg pain separately.

Full-length free-standing lateral spine radiographs (36" cassette) at baseline, 1 year, 2 years, and 3 years follow-up were analyzed using validated software[13,14] (Spineview, ENSAM, Laboratory of Biomechanics, Paris, France). All radiographic measures were performed at a central location based on standard techniques[15] and included: coronal Cobb angles of thoracic and lumbar curves, coronal alignment as measured by the C7 plumbline, thoracic kyphosis (TK, T2–T12; Cobb angle between superior endplate of T2 and inferior endplate of T12), LL (Cobb angle between superior endplate of L1 and superior endplate of S1), SVA (C7 plumbline relative to S1), pelvic tilt (PT), and the mismatch between pelvic incidence and LL (PI-LL). Based on the above radiographic parameters, patients were additionally stratified by the SRS-Schwab adult spinal deformity classification.[16]

Statistical Analyses

Continuous variables were described with the mean and standard deviation. Normality of data was determined using the Shapiro–Wilk test. Comparison of means between the groups included Wilcoxon summed ranked test and Student *t* test with Tukey Honest Significant Difference test to control for type I error where appropriate. Categorical variables were analyzed using Pearson χ^2 test or Fisher exact test where appropriate.

Individual SF36 scores were converted to SF6D utility values[17] based on an externally validated and previously published nonparametric Bayesian model.[18] These were used to calculate QALYs and cumulative QALY gains for the overall cohort at a minimum 2 years postoperative and for a subset of patients that had complete follow-up at 1, 2, and 3 years postoperative. Since this was a trend overtime, they were discounted at the recommended 3.5% per year.[19] The QALY gains were calculated by taking the differences between the postoperative SF6D health-utility states then multiplied by the time spanning the differences; $QALY\ Gained_{3\,\mathrm{yr}} = (SF6D_{3\,\mathrm{yr}} - SF6D_{\mathrm{baseline}}) \times 3$. And for

TABLE 1. Surgical Variables of the Entire Operative Cohort of 258 Patients

Parameter	# (%)
Number of total operative patients	258 (100%)
Posterior approach	255 (98.8%)
Mean levels ± SD (range)	11.1 ± 4.5 (0–19)
Interbody fusion	164 (63.6%)
Mean levels ± SD (range)	2.7 ± 1.7 (1–9)
Anterior lumbar interbody fusion	70 (27.1%)
Posterior interbody fusion	14 (5.4%)
Transforaminal interbody fusion	61 (23.6%)
Lateral interbody procedure	24 (5.4%)
Osteotomy any	171 (66.3%)
Smith–Petersen osteotomy	132 (51.2%)
Pedicle subtraction osteotomy	45 (17.4%)
Corpectomy	2 (0.8%)
Vertebral column resection	9 (3.5%)
Mean operating room time (min, ± SD)	395 ± 133
Mean estimated blood loss (cc, ± SD)	1867 ± 1676
Mean length of stay (days, ± SD)	8.0 ± 4.6
Major complication	112 (43.4%)
Minor complication	127 (49.2%)

SD indicates standard deviation.

example, the 3 year mean QALYs were calculated as follows: $QALY_{3yr} = \frac{SF6D_{1yr}}{1+0.035} + \frac{SF6D_{2yr}}{(1+0.035)^2} + \frac{SF6D_{3yr}}{(1+0.035)^3}$

To account for the nonrandom assignment of operative and nonoperative management between the two cohorts, a propensity score-matching analysis was employed.[20–22] A binominal logistic regression model was created to derive the propensity scores. The target variable was receiving operative or nonoperative treatment with the following covariates for matching: baseline ODI, SRS-22, maximum thoracolumbar/lumbar Cobb angle, PI-LL, C7-S1 SVA, and leg pain NRS score. The matching algorithm used was a 1:1 nearest neighbor matching without replacement, in which each operative case was matched to a unique nonoperative case on the basis of nearest propensity scores. The optimal caliper width for determining the maximum difference between propensity scores to be appropriately matched was calculated by taking 0.2 of the standard deviation of the logit of the propensity scores.[20] Cases were matched to the nearest score first then not to exceed the calculated optimal caliper width. Both unmatched and matched analyses were performed. For the unmatched analyses, all operative and nonoperative patients meeting the study criteria were evaluated. All statistical analyses were conducted using commercially available software (SPSS v22, IBM, Armonk, NY) and the level of significance was set at $P < 0.05$.

RESULTS

Patient Population

Of the available patients eligible for 2-year follow-up (365 operative and 469 nonoperative) 479 had complete baseline and minimum 2-year follow-up data [OP:258 (70.7%), NONOP:221(47.1%)]. There were 406 (84.8%) women and 73 (15.2%) men with a mean age of 54.8 ± 15.7 years. The mean BMI was 26.5 ± 6.1 and mean CCI was 1.2 ± 1.5. Surgical variables and complication rates are presented in Table 1. Of the total 479 patients, 151 (90 OP, 61 NONOP) had complete 1, 2, and 3 year data available for analysis and are reported separately for QALY trending.

Unmatched Mean QALY Analysis

For the entire cohort of 479 patients, OP had significantly greater mean age, BMI, and CCI ($P < 0.05$ for all, Table 2). Regarding the six variables used for propensity score matching, OP had significantly worse mean baseline ODI, SRS total score, leg pain NRS, PI-LL mismatch, and SVA ($P < 0.01$ for all, Table 2) as well as worse baseline mean PT and TK ($P < 0.05$ for both, Table 3). However, they had similar mean baseline maximum coronal Cobb angles and similar baseline distribution of curve locations based on the SRS-Schwab coronal curve classification (Table 2).

OP significantly improved in mean maximum coronal cobb angle, PT, PI-LL, TK, and SVA at minimum 2 years postoperative ($P < 0.05$ for all), whereas NONOP did not have any significant changes in any of the radiographic parameters ($P > 0.05$ for all, Table 3).

OP mean baseline SF6D health utility score was significantly lower NONOP (0.545 ± 0.118 *vs.* 0.657 ± 0.114, respectively, $P < 0.01$, Table 4). Similarly, at minimum 2 years postoperative the mean operative QALY was

TABLE 2. Demographics and Baseline Propensity Score Matching Variables (Marked With **) As Well As Baseline SRS-Schwab Coronal Curve Classification for Both the Matched and Unmatched Entire Operative and Nonoperative Cohorts

	Entire Cohort n = 479			Propensity Score Matched n = 232		
	OP	NONOP	*P* Value	OP	NONOP	*P* Value
Number	258	221		116	116	
Age	57.1 ± 14.5	52 ± 16.6	<0.01*	53.5 ± 16.2	56.3 ± 15	0.24
Female:male	216:42	190:31	0.49	106:10	104:12	0.65
BMI	27.5 ± 6.1	25.5 ± 5.8	<0.01*	26.1 ± 5.3	27.1 ± 6.3	0.31
CCI	1.5 ± 1.6	0.9 ± 1.2	<0.01*	1.1 ± 1.3	1.1 ± 1.3	0.52
Baseline ODI**	42.8 ± 18.8	22.2 ± 15.4	<0.01*	31.2 ± 15.7	30.8 ± 13.5	0.90
Baseline SRS Total**	2.8 ± 0.7	3.6 ± 0.6	<0.01*	3.2 ± 0.5	3.2 ± 0.5	0.87
Baseline Leg pain NRS**	4.3 ± 3.3	2.4 ± 2.8	<0.01*	3.5 ± 3.3	3.1 ± 2.7	0.50
Baseline MAX Thoracolumbar Cobb (deg)**	46.1 ± 19.2	45.1 ± 16.7	0.74	46.6 ± 18.7	46.3 ± 17.8	0.97
Baseline PI-LL (deg)**	15.2 ± 21.1	4.4 ± 16.1	<0.01*	7.5 ± 18.5	7.3 ± 16.4	0.90
Baseline C7-S1 SVA (mm)**	61.7 ± 76.7	17.3 ± 50.5	<0.01*	28 ± 55.8	26.3 ± 52.2	0.82
Preop SRS-Schwab Coronal Curve						
Type N	72 (28.3%)	47 (21.6%)	0.26	27 (23.3%)	22 (19%)	0.52
Type T	15 (5.9%)	17 (7.8%)		10 (8.6%)	7 (6%)	
Type L	93 (36.6%)	78 (35.8%)		38 (32.8%)	48 (41.4%)	
Type D	74 (29.1%)	76 (34.9%)		41 (35.3%)	39 (33.6%)	

SRS-Schwab coronal curve classification: Type D, patients with a double major curve, with each curve greater than 30°; Type L, patients with a lumbar or thoracolumbar major curve of greater than 30° (apical level of T10 or lower); Type N, patients with no coronal curve greater than 30° (i.e. no major coronal deformity); and Type T, patients with a thoracic major curve of greater than 30° (apical level of T9 or higher).

**Represents statistical significance of P < 0.05 between OP and NONOP cohorts.*

***Represents the baseline propensity score matching variables.*

BMI indicates body mass index; CCI, Charlson Comorbidity Index; NRS, numerical rating scale; ODI, Oswestry Disability Index; PI-LL, mismatch between the pelvic incidence (PI) and the lumbar lordosis (LL); SRS, Scoliosis Research Society-22 questionnaire; SVA, sagittal vertical axis.

significantly lower than NONOP (1.278 ± 0.33 *vs.* 1.395 ± 0.374, respectively, $P < 0.01$, Table 4).

For the patients with complete data across 3 years, OP had statistically similar mean age and BMI ($P > 0.05$ for both), but significantly greater CCI ($P < 0.05$, Table 5). For the six variables used for propensity score matching, OP had significantly worse mean baseline ODI and SRS total scores and leg pain NRS ($P < 0.05$ for all), but similar mean maximum coronal Cobb angles, PI-LL, and SVA ($P > 0.05$ for all, Table 5) as well as mean baseline PT and TK ($P < 0.05$ for both, Table 3), as well as similar distribution of curve locations based on the SRS-Schwab coronal curve classification (Table 5).

The mean baseline SF6D health utility score of OP was significantly lower than NONOP (0.566 ± 0.117 *vs.* 0.628 ± 0.105, respectively, $P < 0.01$, Table 4, Figure 1). The mean QALYs at 1, 2, and 3 years postoperative were all statistically similar ($P > 0.05$ for all, Table 4, Figure 1).

Propensity Score Matched Mean QALY Analysis

The matching caliper was 0.285 for the entire cohort of 479 patients and all mean baseline demographic variables, the six mean baseline propensity matching variables as well as all mean baseline radiographic parameters were statistically similar between the groups ($P > 0.05$ for all, Tables 2 and 3).

OP significantly improved in mean maximum coronal Cobb angle, PI-LL and TK at minimum 2 years postoperative ($P < 0.05$ for both), whereas NONOP did not have any significant changes in any of the radiographic parameters ($P > 0.05$ for all, Table 3).

The mean baseline SF6D health utility scores were statistically similar between the groups (OP: 0.609 ± 0.093, NONOP: 0.600 ± 0.091, respectively, $P = 0.64$, Table 4). At minimum 2 years postoperative the mean operative QALY was significantly greater than NONOP (1.377 ± 0.345 *vs.* 1.256 ± 0.286, respectively, $P < 0.01$, Table 4).

As for the subset of patients with complete data across 3 years, the caliper was 0.200 and the mean baseline demographic variables, the six mean baseline propensity matching variables and all mean baseline radiographic parameters were also statistically similar between the groups ($P > 0.05$ for all, Tables 3 and 5).

The mean baseline SF6D health utility scores were statistically similar between the groups (OP: 0.601 ± 0.098, NONOP: 0.602 ± 0.102, respectively, $P = 0.80$, Table 4, Figure 2). OP mean QALYs at 1, 2, and 3 years postoperative were all significantly greater than NONOP ($P < 0.05$ for all, Table 4, Figure 2).

TABLE 3. Baseline and Minimum 2-Year Mean Radiographic Parameters for Both the Matched and Unmatched Operative and Nonoperative Groups for the Entire Cohort

Time	Preop Group	MAX Coronal Cobb	Coronal C7 Alignment (mm)	PT	PI-LL	TK	SVA
Baseline	Operative (n = 258)	46.1 ± 19.2	0.8 ± 48.3	23.5 ± 11.1	15.2 ± 21.1	33.9 ± 17.8	61.7 ± 76.7
	Nonoperative (n = 221)	45.1 ± 16.7	5.4 ± 29.8	19.4 ± 9.9	4.4 ± 16.1	38.6 ± 16.2	17.3 ± 50.5
	P value	0.74	0.29	<0.01*	<0.01*	0.01*	<0.01*
	Operative matched (n = 116)	46.6 ± 18.7	7.9 ± 43.3	20.1 ± 10.8	7.5 ± 18.5	34.3 ± 18	28 ± 55.8
	Nonoperative matched (n = 116)	46.3 ± 17.8	5.4 ± 32.2	20.8 ± 9.7	7.3 ± 16.4	37.2 ± 16.2	26.3 ± 52.2
	P value	0.97	0.89	0.69	0.90	0.16	0.82
Minimum 2 years	Operative (n = 258)	24.2 ± 16.4	4.7 ± 35	21.2 ± 10.1	3.7 ± 15.3	47.5 ± 17.8	30.5 ± 56.4
	Nonoperative (n = 221)	45.2 ± 17.9	8.4 ± 33.7	19.6 ± 10.6	6.2 ± 17.8	37.3 ± 16.8	24.3 ± 54.6
	P value (OP *vs.* NONOP)	<0.01*	0.73	0.08	0.17	<0.01*	0.17
	P value (baseline OP *vs.* min 2 yr OP)	<0.01*	0.39	0.01*	<0.01*	<0.01*	<0.01*
	P value (baseline NONOP *vs.* min 2 yr NONOP)	0.91	0.63	0.78	0.35	0.57	0.33
	Operative matched (n = 116)	24.7 ± 15.7	9.9 ± 33.6	19.8 ± 10.5	1.4 ± 14.6	44.7 ± 17.1	17.2 ± 51.1
	Nonoperative matched (n = 116)	46.1 ± 18.6	5.7 ± 36.8	20.8 ± 10.5	8.5 ± 17.6	37 ± 16.3	28.6 ± 57
	P value	0.42	<0.01*	0.33	<0.01*	<0.01*	0.16
	P value (baseline OP *vs.* min 2 yr OP)	<0.01*	0.45	0.66	<0.01*	<0.01*	0.10
	P value (baseline NONOP *vs.* min 2 yr NONOP)	0.99	0.91	0.90	0.71	0.95	0.96

**Represents statistical significance of P <0.05 between OP and NONOP cohorts.*

PI-LL indicates mismatch between the pelvic incidence (PI) and the lumbar lordosis (LL); PT, pelvic tilt; SVA, sagittal vertical axis (C7-S1); TK, thoracic kyphosis (T2–T12).

TABLE 4. Mean SF6D Baseline Utility Values As Well As the Mean Quality-Adjusted Life Years (QALY) at the Postoperative Time Points for Both Matched and Unmatched Operative (OP) and Nonoperative (NONOP) Cohorts Within the Entire Cohort (TOP) and the Subset of Patients Available for QALY Trends (Bottom)

Entire Cohort (n = 479)	Unmatched			Propensity Score Matched		
	OP	NONOP	*P* Value	OP	NONOP	*P* Value
Number	258	221		116	116	
Baseline SF6D utility	0.545 ± 0.118	0.657 ± 0.114	<0.01*	0.609 ± 0.093	0.600 ± 0.091	0.64
Minimum 2-year QALY	1.278 ± 0.33	1.395 ± 0.374	<0.01*	1.377 ± 0.345	1.256 ± 0.286	<0.01*
Subset With QALY Trend (n = 151)	**Unmatched**			**Propensity Score Matched**		
	OP	**NONOP**	***P* Value**	**OP**	**NONOP**	***P* Value**
Number	90	61		44	44	
Baseline SF6D utility	0.566 ± 0.117	0.628 ± 0.105	<0.01*	0.601 ± 0.098	0.602 ± 0.102	0.80
1-year QALY	0.628 ± 0.096	0.621 ± 0.078	0.61	0.651 ± 0.089	0.61 ± 0.079	0.02*
2-year QALY	1.24 ± 0.183	1.213 ± 0.148	0.21	1.29 ± 0.157	1.189 ± 0.141	<0.01*
3-year QALY	1.827 ± 0.269	1.788 ± 0.21	0.21	1.903 ± 0.235	1.749 ± 0.198	<0.01*

The QALY values presented in the bottom table for the subset of patients for trending were discounted at 3.5% per year.

**Represents statistical significance of P <0.05 between OP and NONOP cohorts.*

TABLE 5. Demographics and Baseline Propensity Score Matching Variables (Marked With **) As Well As Baseline SRS-Schwab Coronal Curve Classification for Both the Matched and Unmatched Operative (OP) and Nonoperative (NONOP) Cohorts Available for QALY Trend Across 1, 2, and 3 Years

	From Entire Cohort n = 151			Propensity Score Matched n = 88		
	OP	NONOP	*P* Value	OP	NONOP	*P* Value
Number	90	61		44	44	
Age	57 ± 13.2	56 ± 15.3	0.81	57 ± 14.8	58 ± 14.2	0.82
Female:male	78:12	54:7	0.74	40:4	39:5	0.73
BMI	26.9 ± 6.4	26.8 ± 6.7	0.59	27 ± 6.7	27.3 ± 7.2	0.89
CCI	1.5 ± 1.5	0.9 ± 1.2	0.01*	1.5 ± 1.5	1 ± 1.3	0.12
Baseline ODI	39.5 ± 18.7	26 ± 16.6	<0.01*	31.8 ± 16.3	30.9 ± 16.3	0.85
Baseline SRS total	2.9 ± 0.7	3.5 ± 0.6	<0.01*	3.3 ± 0.5	3.3 ± 0.6	0.88
Baseline leg pain NRS	4.3 ± 3.4	2.8 ± 3.2	0.02*	3.6 ± 3.3	3.2 ± 3.1	0.80
Baseline MAX thoracolumbar Cobb (°)	48.9 ± 22.1	46 ± 18	0.52	46.8 ± 19.1	47.2 ± 19.8	0.94
Baseline PI-LL (°)	13.4 ± 20.8	8.3 ± 15.2	0.16	10.2 ± 16	8.4 ± 13.9	0.46
Baseline C7-S1 SVA (mm)	56.2 ± 80	35.4 ± 49.6	0.20	38.5 ± 64.8	39.9 ± 50.2	0.93
Preop SRS-Schwab Coronal Curve						
Type N	22 (25%)	13 (21.3%)	0.81	11 (25%)	10 (22.7%)	0.97
Type T	3 (3.4%)	4 (6.6%)		3 (6.8%)	4 (9.1%)	
Type L	34 (38.6%)	24 (39.3%)		18 (40.9%)	17 (38.6%)	
Type D	29 (33%)	20 (32.8%)		12 (27.3%)	13 (29.5%)	

SRS-Schwab coronal curve classification: Type N: patients with no coronal curve greater than 30° (i.e., no major coronal deformity), Type T: patients with a thoracic major curve of greater than 30° (apical level of T9 or higher), Type L: patients with a lumbar or thoracolumbar major curve of greater than 30° (apical level of T10 or lower), and type D: patients with a double major curve, with each curve greater than 30°.

**Represents statistical significance of P <0.05 between OP and NONOP cohorts.*

BMI indicates body mass index; CCI, Charlson Comorbidity Index; NRS, numerical rating scale; ODI, Oswestry Disability Index; PI-LL, mismatch between the pelvic incidence (PI) and the lumbar lordosis (LL); SRS, Scoliosis Research Society-22 questionnaire; SVA, sagittal vertical axis.

Unmatched Cumulative QALYs Gained

OP had a larger QALY gained at minimum 2 years postoperative (0.139 ± 0.253 *vs.* -0.004 ± 0.209, $P < 0.01$). For the subanalysis of the 151 patients with complete 1 to 3 year data, OP had a significantly larger increase in QALYs (from baseline) at 1, 2, and 3 years compared with NONOP (Figure 3): 1 year (0.083 ± 0.115 *vs.* 0.015 ± 0.087, $P < 0.01$), 2 years (0.179 ± 0.245 *vs.* 0.012 ± 0.188, $P < 0.01$), and 3 years (0.254 ± 0.355 *vs.* 0.027 ± 0.257, $P < 0.01$).

Propensity Score Matched Cumulative QALYs Gained

OP had a larger QALY gained at minimum 2 years postoperative (0.112 ± 0.243 *vs.* 0.008 ± 0.195, $P < 0.01$). For the subanalysis of the 151 patients with complete 1 to 3 year

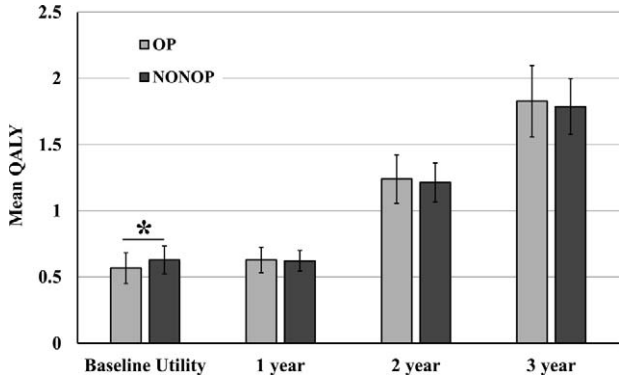

Figure 1. Mean quality-adjusted life years (QALY) for unmatched operative (OP) and nonoperative patients across 3 years postoperative for the subset of patients with complete 1 to 3 year HRQOL (n = 151). (*) represents statistical significance of $P < 0.05$ between the groups.

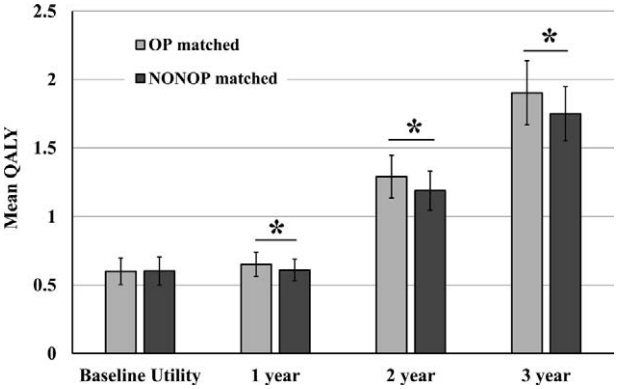

Figure 2. Mean quality-adjusted life years (QALY) for the propensity score matched operative (OP) and nonoperative patients across 3 years postoperative for the subset of patients with complete 1 to 3 year HRQOL (n = 151). (*) represents statistical significance of $P < 0.05$ between the groups.

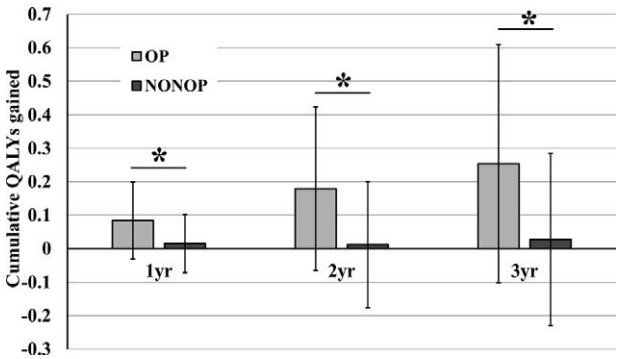

Figure 3. Cumulative QALY gains for unmatched operative (OP) and nonoperative patients across 3 years postoperative for the subset of patients with complete 1 to 3 year HRQOL (n = 151).

data OP had a significantly larger increase in QALYs (from baseline) at 1, 2, and 3 years compared with NONOP (Figure 4): 1 year (0.073 ± 0.121 *vs.* 0.029 ± 0.082, $P = 0.04$), 2 years (0.167 ± 0.232 *vs.* 0.036 ± 0.173, $P < 0.01$), and 3 years (0.238 ± 0.379 *vs.* 0.059 ± 0.258, $P < 0.01$).

DISCUSSION

In the present study, a large number of patients who underwent either operative or nonoperative treatment for adult spinal deformity were evaluated with regard to improvement in quality-adjusted life years. When a 1:1 propensity score match for baseline variables accounting for likelihood of undergoing operative treatment was applied, operative treatment resulted in a significant improvement in mean QALYs over nonoperative treatment at a minimum of 2 years postoperative and even at 1, 2, and 3 years postoperative for the patients with the available data for calculation.

Currently, there is a paucity of QALY data for adult spinal deformity patients and only recently have a few studies been published; most focusing on cost or cost-effectiveness.[1,7–11] One of the first studies to address QALY in ASD was by McCarthy *et al*[8] to determine the incremental cost-effectiveness of operative management. The authors compared 120 surgical and 124 crossover (from initial nonoperative treatment) surgical patients with similar preoperative characteristics from a single institution. QALYs were calculated from the Short Form Health survey and hypothetical nonoperative QALYs were determined based on the crossover group. Total operative costs averaged $125,407 (including readmissions), with average QALYs of 1.93 at 3-year follow-up. This was very close to our value of 1.903 at 3 years of follow-up. The mean hypothetical QALYs without surgery were predicted to be 1.6 after 3 years, whereas in our cohort, the directly calculated nonoperative QALY was higher at 1.749. The QALY calculations were discounted at 3.5% per year for both McCarthy *et al* and ours. Our calculated nonoperative QALYs may be higher since they are directly calculated and account for the QALYs at 1 and 2 years as well. Nonoperative patients may have a larger actual increase in QALY over time than estimated based on baseline characteristics.

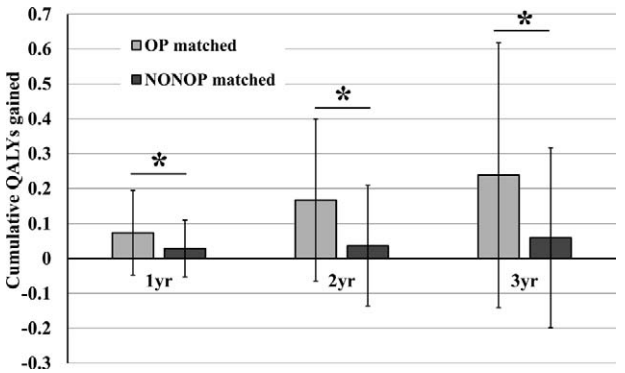

Figure 4. Cumulative QALY gains for matched operative (OP) and nonoperative patients across 3 years postoperative for the subset of patients with complete 1 to 3 year HRQOL (n = 151).

Another cost-effectiveness study was conducted by Terran *et al*[10] in which the authors projected the cost effectiveness of ASD surgery at 5-year follow-up based on 541 patients with 2-year data. The mean cumulative QALYs gained at 1, 2, and predicted 3 years were reported as 0.078, 0.153, and 0.228, respectively. These were calculated by taking the difference in mean health utility values at the given year and by multiplying it by the number of years at the point. For example, the cumulative QALYs gained at 3 years were (3-year health utility value – baseline health utility value) × 3. The calculations are based on the assumption one does not change health states from baseline to the given year of assessment. Our cumulative QALY gains for our overall operative cohort were similar at 0.084, 0.179, 0.258, respectively.

Although the operative groups had statistically greater QALY gains and mean QALYs in the present study, the values were small. This is also reflected in the current literature as well.[8,10,11] Briefly, QALY takes into account both the quantity and quality of life generated by healthcare interventions.[19,23] It places a weight on time based on the patient's health status across that time, usually by each year. Thus, a 1 unit increase in QALYs would mean an added year in perfect health. Moreover, the nonoperative patients gained more QALYs than we anticipated, thus resulting in small differences between the mean QALYs of operative and nonoperative patients. This is largely due to the nature of the QALY analysis. QALY by definition is a health utility value (from 0 to 1) over a certain amount of time, usually yearly.[19] A utility score of 0 indicates death and a score of 1 indicates perfect health. Thus, if one is alive, they simply gain QALYs over time because their score is not 0. QALYs were originally developed in the context of survival analyses to account for both survival and the quality of that survival.[19] However, the mortality for patients with ASD is low and thus, the utility of applying QALY analyses to a population with long survival is questionable. It is widely accepted that ODI and SRS scores are more sensitive to determining outcomes in spine surgery, specifically ASD, than more general health metrics such as SF36 and SF6D. Furthermore, despite the low QALYs gained, operative treatment of ASD has been shown to repeatedly improve

quality of life using these metrics.[24–31] A more sensitive metric may be needed to fully elucidate the benefits of operative treatment of ASD over nonoperative treatment with respect to quality of life and time.

The strengths of the current study include the multicenter design, the largest number of ASD patients to date with direct operative and nonoperative QALY calculations. Furthermore, patients were enrolled from multiple surgeons comprising 11 different sites across the United States, which allows for better generalizability of the results. Another strength of this study is the complete baseline and minimum 2-year follow-up as well as 1, 2, and 3 years follow-up for a subset of patients to assess QALY trends. However, there are a few limitations to this study, one of which includes the retrospective design. Despite the retrospective nature of this study, the data was obtained from a large multicenter prospective database. In addition, a limitation is the relatively lower follow-up rate for nonoperative compared with operative patients, a challenge that has been reported in multiple previous studies.[28,32–35] Moreover, loss to follow-up in the nonoperative cohort may overestimate the reduced QALYs in this cohort, thus exaggerating the potential benefits of surgery. Last, limitation is the lack of cost data associated with the QALYs calculated for this large cohort. Thus, cost-effectiveness could not be directly assessed as of yet. However, our group is currently retrospectively collecting this data for the reported patients and will be performing an extensive cost-effectiveness analysis of operative management for ASD patients compared with nonoperative management in the near future.

CONCLUSION

The operative treatment of adult spinal deformity results in significant higher mean QALYs and larger QALYs gained and at minimum 2 years postop as well as at the 1-, 2-, and 3-year time points compared with nonoperative management.

➢ Key Points

- ❑ Operative patients had significantly worse baseline SF6D health utility states than the nonoperative patients.
- ❑ Operative patients had significantly greater mean QALY at minimum 2 years postoperative and at 1, 2, and 3 years.
- ❑ Operative patients had significantly greater QALYs gained at minimum 2 years postoperative and at 1, 2, and 3 years.

References

1. McCarthy IM, Hostin RA, O'Brien MF, et al. Analysis of the direct cost of surgery for four diagnostic categories of adult spinal deformity. *Spine J* 2013;13:1843–8.
2. Carter OD, Haynes SG. Prevalence rates for scoliosis in US adults: results from the first National Health and Nutrition Examination Survey. *Int J Epidemiol* 1987;16:537–44.
3. Francis RS. Scoliosis screening of 3,000 college-aged women. The Utah Study–phase 2. *Phys Ther* 1988;68:1513–6.
4. Kostuik JP, Bentivoglio J. The incidence of low-back pain in adult scoliosis. *Spine (Phila Pa 1976)* 1981;6:268–73.
5. Perennou D, Marcelli C, Herisson C, et al. Adult lumbar scoliosis. Epidemiologic aspects in a low-back pain population. *Spine (Phila Pa 1976)* 1994;19:123–8.
6. Schwab F, Dubey A, Gamez L, et al. Adult scoliosis: prevalence, SF-36, and nutritional parameters in an elderly volunteer population. *Spine (Phila Pa 1976)* 2005;30:1082–5.
7. McCarthy I, Hostin R, O'Brien M, et al. Health economic analysis of adult deformity surgery. *Neurosurg Clin North Am* 2013;24: 293–304.
8. McCarthy I, O'Brien M, Ames C, et al. Incremental cost-effectiveness of adult spinal deformity surgery: observed quality-adjusted life years with surgery compared with predicted quality-adjusted life years without surgery. *Neurosurg Focus* 2014;36:E3.
9. McCarthy IM, Hostin RA, Ames CP, et al. Total hospital costs of surgical treatment for adult spinal deformity: an extended follow-up study. *Spine J* 2014;14:2326–33.
10. Terran J, McHugh BJ, Fischer CR, et al. Surgical treatment for adult spinal deformity: projected cost effectiveness at 5-year follow-up. *Ochsner J* 2014;14:14–22.
11. Fischer CR, Terran J, Lonner B, et al. Factors predicting cost-effectiveness of adult spinal deformity surgery at 2 years. *Spine Deform* 2014;2:415–22.
12. Charlson M, Szatrowski TP, Peterson J, et al. Validation of a combined comorbidity index. *J Clin Epidemiol* 1994;47:1245–51.
13. Champain S, Benchikh K, Nogier A, et al. Validation of new clinical quantitative analysis software applicable in spine orthopaedic studies. *Eur Spine J* 2006;15:982–91.
14. Rillardon L, Levassor N, Guigui P, et al. Validation of a tool to measure pelvic and spinal parameters of sagittal balance. *Rev Chir Orthop Reparatrice Appar Mot* 2003;89:218–27.
15. O'Brien MF, Kuklo TR, Blanke K, editors. et al. *Spinal Deformity Study Group Radiographic Measurement Manualed*. Memphis, TN: Medtronic Sofamor Danek; 2005.
16. Schwab F, Ungar B, Blondel B, et al. Scoliosis Research Society-Schwab adult spinal deformity classification: a validation study. *Spine (Phila Pa 1976)* 2012;37:1077–82.
17. Craig BM, Pickard AS, Stolk E, et al. US valuation of the SF-6D. *Med Decis Making* 2013;33:793–803.
18. Kharroubi S, Brazier JE, O'Hagan A. Modelling covariates for the SF-6D standard gamble health state preference data using a non-parametric Bayesian method. *Soc Sci Med* 2007;64:1242–52.
19. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. *Br Med Bull* 2010;96:5–21.
20. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat* 2011;10:150–61.
21. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;46:399–424.
22. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;17:2265–81.
23. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. *Value Health* 2009;12 (suppl 1):S5–9.
24. Hyun SJ, Rhim SC. Clinical outcomes and complications after pedicle subtraction osteotomy for fixed sagittal imbalance patients: a long-term follow-up data. *J Korean Neurosurg Soc* 2010;47:95–101.
25. Ikenaga M, Shikata J, Takemoto M, et al. Clinical outcomes and complications after pedicle subtraction osteotomy for correction of thoracolumbar kyphosis. *J Neurosurg Spine* 2007;6:330–6.
26. Liu S, Schwab F, Smith JS, et al. Likelihood of reaching minimal clinically important difference in adult spinal deformity: a

comparison of operative and nonoperative treatment. *Ochsner J* 2014;14:67–77.
27. Smith JS, Klineberg E, Schwab F, 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 non-operative treatment. *Spine (Phila Pa 1976)* 2013;38:1663–71.
28. Smith JS, Shaffrey CI, Berven S, et al. 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.
29. Smith JS, Shaffrey CI, Glassman SD, et al. Clinical and radiographic parameters that distinguish between the best and worst outcomes of scoliosis surgery for adults. *Eur Spine J* 2013;22:402–10.
30. Terran J, Schwab F, Shaffrey CI, et al. The SRS-Schwab adult spinal deformity classification: assessment and clinical correlations based on a prospective operative and nonoperative cohort. *Neurosurgery* 2013;73:559–68.
31. Yang BP, Ondra SL, Chen LA, et al. Clinical and radiographic outcomes of thoracic and lumbar pedicle subtraction osteotomy for fixed sagittal imbalance. *J Neurosurg Spine* 2006;5:9–17.
32. Bridwell KH, Baldus C, Berven S, 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;35:1849–54.
33. Bridwell KH, Glassman S, Horton W, 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;34:2171–8.
34. Smith JS, Shaffrey CI, Berven S, et al. Improvement of back pain with operative and nonoperative treatment in adults with scoliosis. *Neurosurgery* 2009;65:86–93.
35. Smith JS, Shaffrey CI, Glassman SD, et al. Risk-benefit assessment of surgery for adult scoliosis: an analysis based on patient age. *Spine (Phila Pa 1976)* 2011;36:817–24.