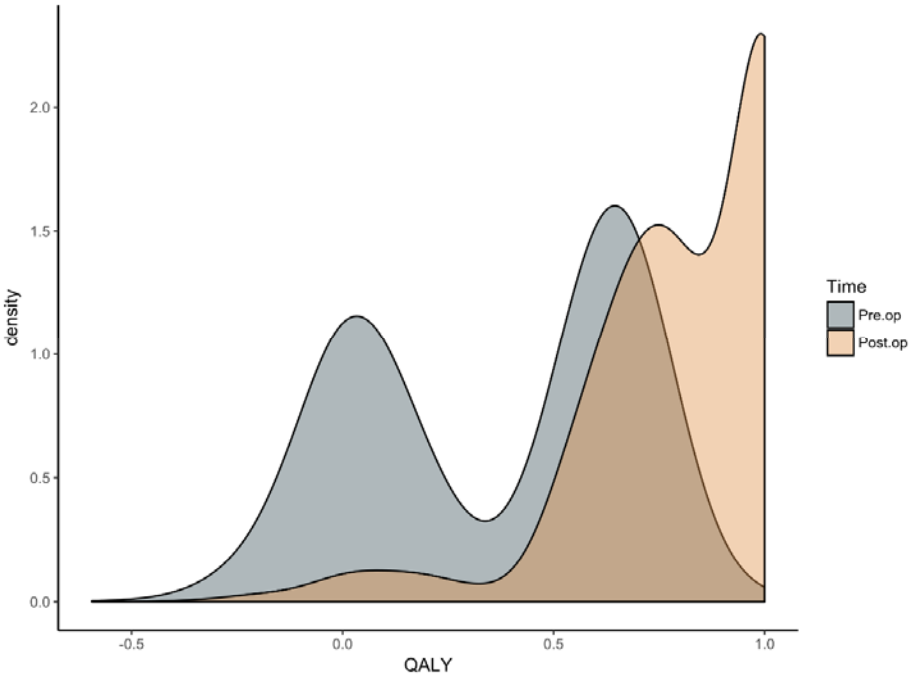
Development of Deployable Predictive Models for MCID of 2 year Outcomes across All Commonly Used HRQOL Instruments in Adult Spinal Deformity Surgery: Results in 570 Patients from 17 Hospitals

Miquel Serra-Burriel
Michael P. Kelly
Justin S Smith
Jeffrey L Gum
Ferran Pellisé
Ahmet Alanay
Emre R Acaroglu
Francisco Javier Sánchez Pérez-Grueso
Frank S. Kleinstück
Ibrahim Obeid
Virginie Lafage
Frank J. Schwab
Christopher I. Shaffrey
Douglas C. Burton
Shay Bess
Christopher P. Ames
European Spine Study Group
International Spine Study Group

- ASD surgery is costly with variable outcomes; in some series only 50% of patients achieve MCID improvements.
- 1 & 2-year outcomes after ASD surgery present highly inherent variance
- Self-reported HRQOL are not perfect measures, but are the best up-to-date
- Shared-decision making: moving from *ambiguity to uncertainty*
- Predictive analytics may accurately model HRQOL improvements after ASD surgery.
- Predictive modeling may be useful in shared-decision making and surgical planning.

Self-reported HRQoL measures – EQ5D

- PROMS present inherent measurement error, 10%
- What matters most, sample size or data quality?
- Comparison with Hip replacement, Knee replacement and Groin hernia repair (Gutacker & Street 2017)



Hip Replacement – QALY distribution

Qual Life Res (2017) 26:2497–2505

Procedure	#groups	Development sample		Test sample		Reduced model	
		$adj\text{-}R^2$	RMSE	$adj\text{-}R^2$	RMSE	$adj\text{-}R^2$	RMSE
Hip replacement	55	14.3%	0.228	12.8%	0.218	1.5%	0.244
Knee replacement	59	19.4%	0.231	18.8%	0.224	2.1%	0.255
Groin hernia repair	60	27.0%	0.161	28.1%	0.158	1.3%	0.188

Development sample: April 2009 to March 2015. Test sample: April 2015 to March 2016. Reduced model only considers age, sex and symptom period for grouping and is estimated and tested on the development sample. R^2 is adjusted for number of predictor variables, i.e. groups

Replication, 28 controls, comorbidities and preop status

N=34,777, Training (N=26,084), Testing (N=8,693)

RMSE	Rsquared	MAE
0.2157674	0.1932373	0.1613225

30.2 – 82.5% precision in our models depending upon time frame & outcome

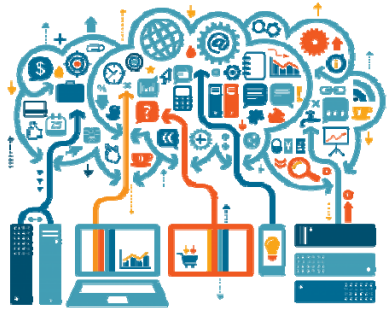
The objective of this study was to model HRQOL overall improvement, including the likelihood of achieving clinically important improvement, at 2 years postoperatively.

Methods

Retrospective analysis of two independent, prospective, multi-center ASD databases with identical fixed data fields

- European Spine Study Group (ESSG) Database
- International Spine Study Group (ISSG) Database

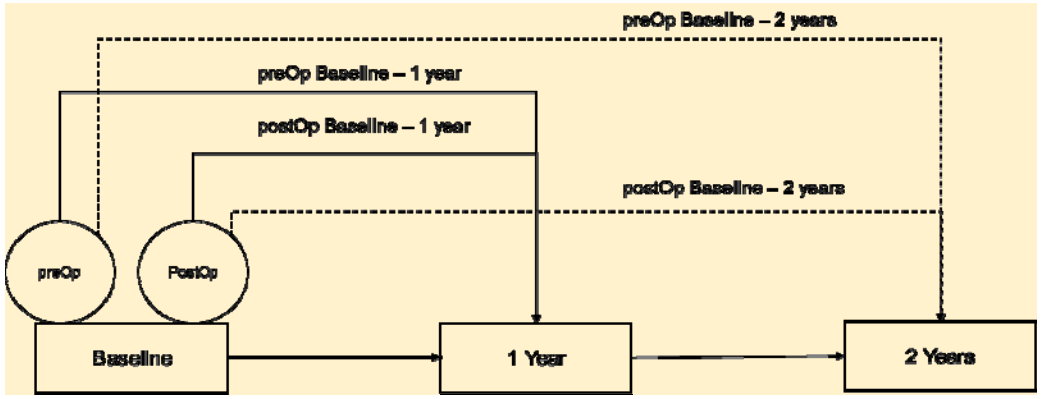
Predictive modeling (Machine Learning)



75 variables were used in the training of the models including:

a) patient characteristics
b) Surgical characteristics
c) Site-Fixed effects

8 different prediction algorithms were trained with 3-time horizons: baseline to 1yr-BL, 2yr-BL and 1yr-2yr:



- External validation was accomplished via an 80/20 data split for training and testing each model, respectively.
- 5-Fold cross validation within the training sample was performed.

Accuracy was measured as the mean average error (MAE; smaller is better) and R^2 values.

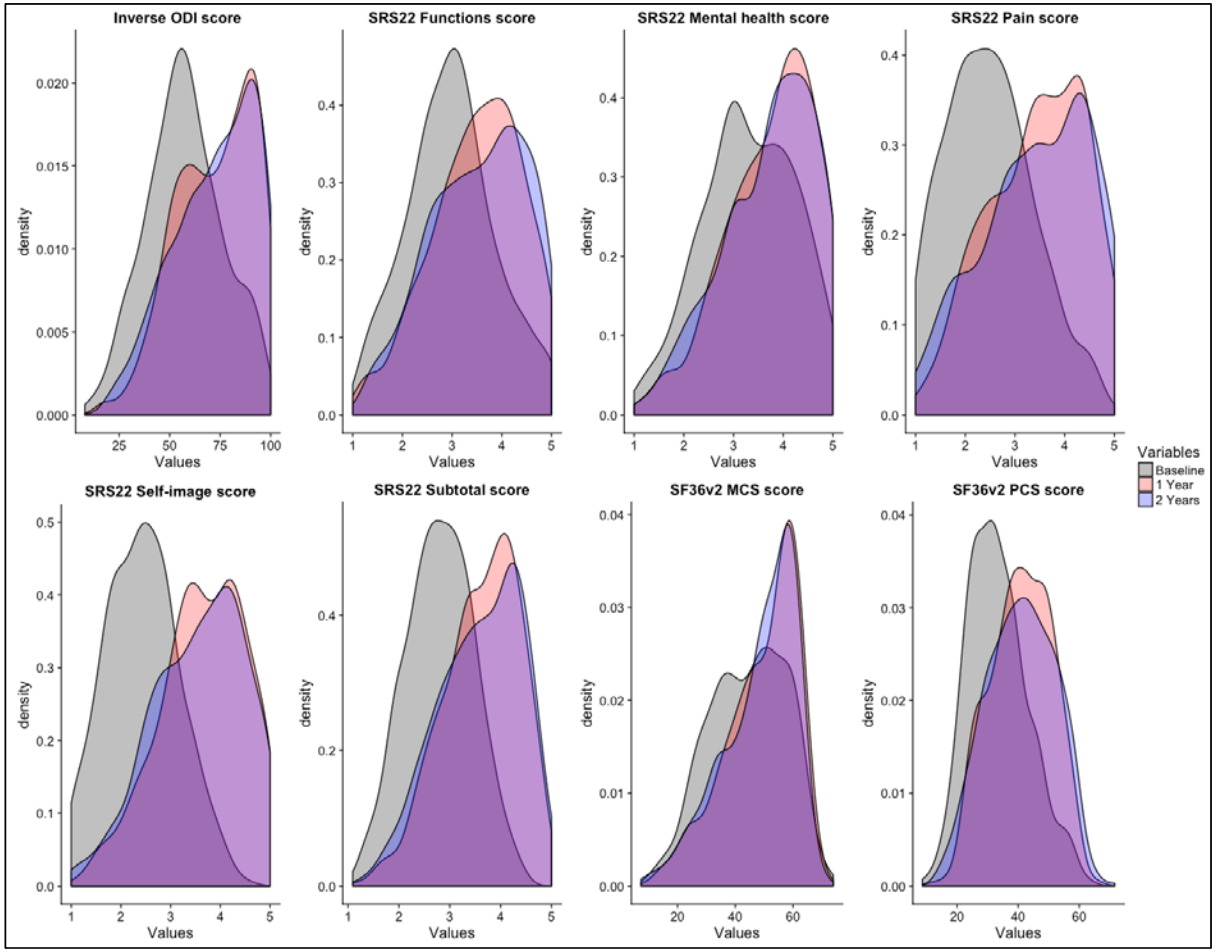
Inclusion criteria:

ASD ≥ 18 years
HRQoL scores (ODI, SRS22,SF36v2)
Surgically treated

1612 surgical patients:
24 hospitals
5 countries
57 surgeons
730.5 days mean FU

743 (46.1%) excluded since they were not yet eligible for 2-year follow-up due to on-going database enrollment

570 met inclusion criteria



Kernel distribution of Scores across the 8 outcomes at Baseline, 1 Year and 2 Years.
The shift to the right represents the population improvement.

Variable	Mean or %	SD	Min	Max
Pre-operative Baseline				
Gender (% women)	79			
Height (cm)	162.55	9.61	104.5	195.6
Weight (kg)	70.31	16.03	40	158.8
Age (years)	56.75	16.32	18.05	86.42
Sagittal alignment (SVA, cm)	55.64	69.55	-83.73	314.09
Coronal alignment (GCA, cm)	31.24	31.4	0	288.95
Major curve Cobb angle (°)	40.27	21	0	123.7
Pelvic tilt (°)	23.28	11.03	-12.66	67.9
1-Year Post-operative				
Sagittal alignment (SVA, cm)	23.83	51.38	-114.08	181.03
Coronal alignment (GCA, cm)	4.98	30.23	-162.17	113.25
Major curve Cobb angle (°)	9.09	23.43	-77.5	100.2
Pelvic tilt (°)	20.83	9.6	-15.29	61.13
2-Years Post-operative				
Sagittal alignment (SVA, mm)	30.36	54.23	-109.2	237.1
Coronal alignment (GCA, mm)	23.37	19.57	0	114.8
Major curve Cobb angle (°)	20.59	16.35	0	94.5
Pelvic tilt (°)	20.87	10.11	-6.3	61.87

Variable	Mean or %	SD	Min	Max
Previous spine surgery (%)	41.6			
Any major complication (%)	36.2			
Total surgical time (min)	345.74	134.47	43	802
Estimated blood loss (cc)	1660.16	1465.51	20	12200
Number of fused vertebral levels	10.7	4.21	0	23
Use of pelvic fixation (%)	59.5			
Use of interbody fusion (%)	54.4			
Use of Smith-Petersen osteotomy (%)	50.0			
Use of 3-column osteotomy (%)	20.8			
Length of hospitalization (days)	8.88	6.49	0	83

Variable	Mean or %	SD	Min	Max
Pre-operative Baseline				
Inverse ODI score[2]	57.55	18.89	8	100
SRS-22r function score	2.98	0.88	1	5
SRS-22r mental health score	3.31	0.93	1	5
SRS-22r pain score	2.49	0.88	1	5
SRS-22r self-image score	2.43	0.73	1	4.6
SRS-22r subtotal score	2.8	0.66	1.09	4.5
SF36v2 MCS score	44.53	13.19	10.59	73.53
SF36v2 PCS score	33.38	9.85	8.65	63.58
1-Year Post-operative				
Inverse ODI score[2]	72.44	18.94	16	100
SRS-22r function score	3.5	0.9	1	5
SRS-22r mental health score	3.75	0.88	1	5
SRS-22r pain score	3.45	0.96	1	5
SRS-22r self-image score	3.64	0.86	1	5
SRS-22r subtotal score	3.63	0.74	1.14	5
SF36v2 MCS score	49.34	12.32	10.48	73.95
SF36v2 PCS score	40.98	10.03	11.47	66.45
2-Years Post-operative				
Inverse ODI score[2]	72.51	19.98	17.78	100
SRS-22r function score	3.53	0.95	1.2	5
SRS-22r mental health score	3.72	0.92	1	5
SRS-22r pain score	3.44	1.06	1	5
SRS-22r self-image score	3.57	0.91	1	5
SRS-22r subtotal score	3.6	0.81	1.23	5
SF36v2 MCS score	48.57	12.43	7.32	70.87
SF36v2 PCS score	40.95	11.06	13.01	71.61

- Baseline HRQL scores are the main predictors of the outcomes after surgery (65%), the present is the best predictor of the future.

- Across all analyses there are some salient features:
  - baseline physical status is a determinant of future HRQL; patients with worse initial state present significant greater relative improvements than patients with better initial status.
  - Prior spine surgery is also a big predictor of both HRQL improvement and major complications incidence.
  - The additional relative predictive gain of surgical factors is limited; i.e. the number of SPOs compared to baseline status only presents an adjusted 2% relative importance.
  - Arthritis, hypertension and depression are comorbidities that have a greater effect on HRQL at 2 years.

Names	Importance
SRS22 subtotal score Baseline	**12.57%**
ODI score Baseline	**10.32%**
SF36v2 MCS score Baseline	**9.26%**
SRS22 MH score Baseline	**8.31%**
SF36v2 PCS score Baseline	**7.74%**
SRS22 function score Baseline	**7.33%**
SRS22 pain score Baseline	**5.73%**
Age Baseline	4.44%
SRS22 SI score Baseline	**3.82%**
Level of physical labor Null	3.23%
Prior Spine Surgery Yes	3.11%
Pelvic Tilt Baseline	2.45%
Hypertension Baseline	2.35%
Arthritis Baseline Yes	2.31%
Fused vertebrae	2.26%
Height cm Baseline	2.26%
Sagittal Balance Baseline	2.25%
Number of levels between UIV and LIV	2.22%
Major Curve Cobb Angle Baseline	2.12%
Depression Baseline Yes	2.00%
Weight kgs Baseline	1.97%
SPOs	1.95%

Aggregated Relative Variable Importance

*It is important to not interpret this variables as having a causal effect, due to the retrospective nature of the data

Models with the lowest MAE for each of the 5-time points were selected; ultimately the model had 82.4% predictive power. Patients with lower enrollment HRQOL were likely to appreciate the greatest improvements in HRQOL at 2-yr followup. Addition of surgeon and site to preoperative data increased the predictive power 1.8%. Site and surgeon fixed-effects played a crucial role in explaining outcome variance.

	HRQL instrument	Baseline Score	Baseline Probabilities	Depression and Hypertension	10% improve in physical HRQL scores	Waiting 5 years with a reduction of 10 % HRQL	Waiting 10 years with a reduction of 20 % HRQL	Post-op SVA 0mm	Post-op SVA 50mm	Post-op SVA 100mm	Number of Levels from 10 to 15	Range of variation across options
(1) Overall improvement	ODI	68	54.4%	55.1%	43.5%	67.6%	72.6%	54.4%	54.4%	54.4%	54.4%	29.1%
(2) Improvement over MCID	ODI	68	38.4%	39.1%	28.4%	52.0%	57.7%	38.4%	38.4%	38.4%	38.4%	29.3%
(1) Overall improvement	SRS22 function	3.2	65.0%	66.1%	51.4%	76.4%	83.5%	65.0%	65.0%	65.0%	65.0%	32.1%
(2) Improvement over MCID	SRS function	3.2	35.4%	36.5%	23.4%	48.3%	58.4%	35.4%	35.4%	35.4%	35.4%	35.0%
(1) Overall improvement	SRS22 MH	2.4	80.6%	80.6%	82.0%	86.0%	90.3%	80.6%	80.6%	80.6%	78.6%	11.7%
(2) Improvement over MCID	SRS22 MH	2.4	64.7%	64.7%	66.6%	72.4%	79.2%	64.7%	64.7%	64.7%	62.0%	17.2%
(1) Overall improvement	SRS22 pain	2	61.4%	63.7%	67.1%	76.2%	82.6%	61.4%	61.4%	61.4%	62.2%	21.2%
(2) Improvement over MCID	SRS22 pain	2	46.2%	48.6%	52.3%	62.9%	71.0%	46.2%	46.2%	46.2%	47.0%	24.8%
(1) Overall improvement	SRS22 SI	2.4	81.5%	81.5%	83.7%	84.3%	86.9%	81.5%	81.5%	81.5%	81.5%	5.4%
(2) Improvement over MCID	SRS22 SI	2.4	34.6%	34.6%	37.8%	38.8%	43.3%	34.6%	34.6%	34.6%	34.6%	8.7%
(1) Overall improvement	SRS22 subtotal	2.38	69.8%	72.1%	74.7%	84.8%	91.7%	69.8%	69.8%	69.8%	70.7%	21.9%
(2) Improvement over MCID	SRS22 subtotal	2.38	45.8%	48.4%	51.7%	65.6%	77.7%	45.8%	45.8%	45.8%	46.8%	31.9%
(1) Overall improvement	SF36v2 MCS	22.18	95.6%	95.4%	95.6%	97.0%	98.0%	95.6%	95.6%	95.6%	95.6%	2.6%
(2) Improvement over MCID	SF36v2 MCS	22.18	81.5%	81.0%	81.5%	85.9%	89.5%	81.5%	81.5%	81.5%	81.5%	8.5%
(1) Overall improvement	SF36v2 PCS	39.66	44.3%	45.3%	34.6%	60.4%	63.4%	44.3%	44.3%	44.3%	43.5%	28.8%
(2) Improvement over MCID	SF36v2 PCS	39.66	17.4%	18.0%	11.6%	29.7%	32.4%	17.4%	17.4%	17.4%	16.9%	20.8%

We present an accurate and consistent way of predicting outcome scores for ASD surgery in the largest-to-date prospective operative multicenter cohort with 2-year follow-up.

This study has significant clinical implications for shared-decision making, surgical planning and postoperative counseling.

Surgeon and site were important components of the model, explaining variance in predicted 2-yr HRQOL.

Miquel Serra-Burriel:	None
Michael P. Kelly	Grants from AOSpine, grants from DePuy Synthes, grants from DePuy Synthes, outside the submitted work.
Justin S Smith:	Grants from DePuy Synthes/ISSG, during the conduct of the study; personal fees from Zimmer Biomet, personal fees from Nuvasive, personal fees from K2M, personal fees from AlloSource, from Cerapedics, grants from NREF, grants from AOSpine, from DePuy Synthes/ISSG, outside the submitted work.
Jeffrey L Gum:	Grants from Depuy Synthes, personal fees from Acuity, personal fees from Alphatec Spine, personal fees from DePuy Synthes, personal fees from K2M, personal fees from Medtronic, personal fees from Nuvasive, personal fees from Stryker Spine, outside the submitted work
Ferran Pellisé:	Grants from Depuy Synthes, grants from Fondo de Investigaciones Sanitarias, during the conduct of the study; personal fees from Depuy Synthes, grants from Medtronic, grants from Zimmer-Biomet, outside the submitted work.
Ahmet Alanay:	Grants from Depuy Synthes, during the conduct of the study; grants from DePuy & Synthes, grants from outside the submitted work.
Emre R Acaroglu:	Grants from Depuy Synthes, during the conduct of the study; grants from Medtronic, personal fees from AOSpine, outside the submitted work.
Francisco JS Pérez-Grueso:	Grants and other from Depuy Synthes, other from K2M, outside the submitted work.
Frank S. Kleinstück:	Grants from Depuy Synthes, during the conduct of the study.
Ibrahim Obeid:	Grants from Depuy Synthes, during the conduct of the study; personal fees from Depuy Synthes, personal fees from Medtronic, personal fees from Alphatec spine, personal fees from Spineart, personal fees from Clariance, outside the submitted work.
Virginie Lafage:	Grants from SRS, NASS, grants from DePuy Spine, grants from NuVasive, grants from Stryker, grants from K2M, personal fees from Depuy Synthes, personal fees from AO Spine, personal fees from K2M, outside the submitted work.
Frank J. Schwab:	Grants from DePuy Spine, grants from Stryker, grants from K2M, grants from NuVasive, personal fees from Medicrea, personal fees from Zimmer-Biomet, personal fees from NuVasive, personal fees from K2M, personal fees from MSD, other from Nemaris INC, outside the submitted work.
Christopher I. Shaffrey:	Grants from ISSG, during the conduct of the study; personal fees from Medtronic, personal fees from NuVasive, personal fees from Zimmer Biomet, outside the submitted work.
Douglas C. Burton:	Grants, personal fees and other from Depuy Synthes, nonfinancial support from ISSG, non-financial support from SRS, non-financial support from University of Kansas Physicians, other from Bioventus, other from Pfizer, outside the submitted work.
Shay Bess:	Grants from Depuy Synthes, grants from K2M, during the conduct of the study; grants from NuVasive, grants from Medtronic, grants from Stryker Spine, grants from Biomet Spine, grants from Orthofix, personal fees from K2M outside the submitted work.
Christopher P. Ames:	Personal fees from DePuy Synthes, Stryker, Biomet Spine, Nuvasive, Next Orthosurgical, DePuy Synthes, Medtronic, Stryker, Medicrea, K2M, UCSF, outside the submitted work.
ESSG:	Funding support from DePuy Synthes and Medtronic. Additional support was provided through Project PI16/01283, funded by Instituto de Salud Carlos III and co-funded by EU (ERDF/ESF).
ISSG:	Funding support from DePuy Synthes, K2M, Nuvasive, Orthofix, and Zimmer Biomet.