



Published in final edited form as:

Spine (Phila Pa 1976). 2023 January 01; 48(1): 21–28. doi:10.1097/BRS.0000000000004416.

Validation of Adult Spinal Deformity Surgical Outcome Prediction Tools in Adult Symptomatic Lumbar Scoliosis

James P. Wondra II, BS^{1,*}, Michael P. Kelly, MD^{2,*}, Jacob Greenberg, MD³, Elizabeth L. Yanik, PhD¹, Christopher Ames, MD⁴, Ferran Pellise, MD⁵, Alba Vila-Casademunt, MSc⁵, Justin S. Smith, MD, PhD⁶, Shay Bess, MD⁷, Christopher Shaffrey, MD⁸, Lawrence G. Lenke, MD⁹, Miquel Serra-Burriel, PhD¹⁰, Keith Bridwell, MD¹

¹Department of Orthopedic Surgery, Washington University School of Medicine, St. Louis, Missouri.

²Department of Orthopaedic Surgery, Rady Children's Hospital, University of California, San Diego, San Diego, CA

³Department of Neurological Surgery, Washington University School of Medicine, St. Louis, Missouri

⁴Department of Neurosurgery, University of California, San Francisco, California. Etc.

⁵Spine Unit, Hospital Vall d'Hebron, Barcelona Spain

⁶Department of Neurological Surgery, University of Virginia, Charlottesville, VA

⁷Denver International Spine Center, Denver, Colorado

⁸Department of Neurological Surgery, Duke University, Durham, NC

⁹Och Spine Hospital, Columbia University College of Physicians and Surgeons, New York, NY

¹⁰Center for Research in Health and Economics, Universitat Pompeu Fabra, Barcelona, Spain

Abstract

Study Design: Post-hoc analysis

Objective: Advances in machine learning have led to tools offering individualized outcome predictions for adult spinal deformity (ASD). Our objective is to examine the properties of these ASD models in a cohort of adult symptomatic lumbar scoliosis (ASLS) patients.

Summary of Background Data: Machine-learning algorithms produce patient-specific probabilities of outcomes, including major complication, reoperation, and readmission in ASD. External validation of these models is needed.

Corresponding Author's name and current institution: Michael P. Kelly: Department of Orthopaedic Surgery, Rady Children's Hospital, University of California, San Diego, San Diego, CA, mkelly5@rchsd.org.

*These authors contributed equally to this work.

Author Contributions:

Conception and design: all authors. Acquisition of data: Kelly, Smith, Shaffrey, Bess, Lenke, Bridwell Analysis and interpretation of data: Wondra, Kelly, Greenberg, Yanik, Ames, Pellise, Smith. Drafting the article: Wondra, Kelly. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Kelly

Methods: 39 predictive factors (12 demographic, 9 radiographic, 4 health-related quality of life, 14 surgical) were retrieved and entered into web-based prediction models for major complications (MC), unplanned reoperation (RO), and hospital readmission (RA). Calculated probabilities were compared with actual event rates. Discrimination and calibration were analyzed using receiver operative characteristic area under the curve (ROC/AUC where 0.5=chance, 1=perfect) and calibration curves (Brier scores, where 0.25=chance, 0=perfect). 95% confidence intervals are reported.

Results: 169 of 187 (90%) surgical patients completed 2-year follow up. The observed rate of major complications was 41.4% with model predictions ranging from 13%–68% (mean 38.7%). Reoperation was 20.7% with model predictions ranging from 9%–54% (mean 30.1%). Hospital readmission was 17.2% with model predictions ranging from 13%–50% (mean 28.5%). Model classification for all three outcome measures was better than chance for all (AUC = MC 0.6 (0.5–0.7), RA 0.6 (0.5–0.7), RO 0.6 (0.5–0.7)). Calibration was better than chance for all, though best for RA and RO (Brier Score = MC 0.22, RA 0.16, RO 0.17).

Conclusions: ASD prediction models for major complication, readmission, and reoperation performed better than chance in a cohort of adult lumbar scoliosis patients, though the homogeneity of ASLS affected calibration and accuracy. Optimization of models require samples with the breadth of outcomes (0–100%), supporting the need for continued data collection as personalized prediction models may improve decision-making for the patient and surgeon alike.

Keywords

3 to 6 words or phrases: adult; lumbar; outcomes; predictive modeling; scoliosis; spine deformity; surgery

Introduction:

Adult symptomatic lumbar scoliosis (ASLS) is a subset of adult spinal deformity (ASD) with increasing prevalence in our aging population.^{1–3} Surgery is more effective than nonoperative treatment for patients seeking meaningful improvement in health-related quality of life (HRQoL).^{4–7} However, not all patients benefit, and surgery entails significant expense and risk for adverse events.^{8–10} With a growing emphasis on value-driven healthcare and an increasing volume of ASD surgeries, accurately predicting which patients are most likely to benefit from surgery is critical.^{11–13}

The predictive power of patient centered decision aids is rapidly improving due to advances in artificial intelligence (AI), including machine learning (ML).¹⁴ ML uses empirical patient data such as demographic data, patient reported outcome measures (PROMs), and surgical outcomes data to create mathematical models that describe the complex relationships between these variables.¹⁵ When used with large datasets of patient data, ML-modeling predictions can be developed into patient centered tools to help inform surgical counseling. Predictive modeling in healthcare is in its infancy and has the potential to revolutionize personalized care across our value-driven healthcare economy.^{15–18}

Decision aid use in spine surgery has progressed from general surgical risk calculators to spine-specific tools and has the potential to offer patient personalized models.^{14,19–24} The

International Spine Study Group (ISSG) and European Spine Study Group (ESSG) merged data to create a voluminous surgical ASD dataset amenable to ML techniques. Models to predict major complication (MC), readmission (RA), reoperation (RO), patient-reported outcome measures (PROMs), and costs were created.^{25–27} In each case the model performed better than traditional statistical methods in terms of predicting the probability of these events. Development of these models required vast amounts of data, however, and manual entry in the clinic would be unwieldy. As a result, parsimonious “simple” models were created to minimize data requirements with minimal effect on prediction accuracy. This study sought to examine validity of the “simple” models to predict MC, RO, and RA at up to two years after surgery in a cohort of ASLS patients.

Methods:

Study Population

This is a secondary analysis of data collected during the ASLS-1 study.⁴ The study was conducted at nine centers in North America and included randomized and observational patient cohorts. All sites obtained IRB approval and the study was registered with [ClinicalTrials.gov \(NCT00854828\)](https://clinicaltrials.gov/ct2/show/study/NCT00854828). Eligible patients were 40–80 years old with ASLS, defined as a lumbar curve with a coronal Cobb measurement $\geq 30^\circ$ and ODI score of 20 or SRS-22 score ≥ 4.0 in pain, function, and/or self-image domains. Demographic data were obtained by standardized case report forms and all patients were deemed surgical candidates by the treating surgeon. Data were reviewed for completeness and accuracy by the study coordinators and a monitoring board. Patients with prior spinal fusion or multilevel decompression surgery were excluded. Enrollment began in April 2010 and ended July 2014. This is an as-treated analysis of all patients (randomized and observational) treated with surgery during the conduct of the study. Descriptive and surgical data were collected and compared with data from the ISSG-ESSG development dataset.²⁷ Unpaired t-tests and Chi-square analyses were performed as appropriate. No correction for multiple comparisons were made and statistical significance was defined as $p < 0.05$.

Predictive Modeling

The previously created ASD predictive models were accessed and the “simple model” option was selected to predict probabilities of major complication, readmission, and reoperation for each individual patient.^{26,27} The “simple model” was created to make parsimonious models and reduce data requirements while preserving prediction accuracy. Thirty-nine predictive factors were input for each ASLS patient, including: 12 demographic factors, 9 radiographic factors, 4 HRQoL factors, and 14 surgical factors. Individual factors are summarized in Table 1. HRQoL factors were garnered from patient questionnaires, including two questions from the Oswestry Disability Index (ODI) regarding walking and homemaking, one Scoliosis Research Society (SRS)-22r question regarding medication usage, and one Short Form (SF)-36 health survey question regarding daily activities.^{29–31} Required model parameters not included in the ASLS-1 study were American Society of Anesthesiologists (ASA) grade, heel walk, toe walk, and leg length discrepancy. ASA was estimated from patient comorbidities as suggested by *Mannion* et al.³² Heel-walk

and toe-walk were estimated from a lower extremity motor deficit screening. Leg length discrepancies were entered as zero.

Patient specific probabilities for sustaining MC, RA, or RO were generated and recorded. MC, RA, and RO were predicted and followed over a 2-year postoperative timeline.

Validation and Statistical Analysis

Model generated predictions were compared to actual events in the surgically treated ASLS-1 cohort. Model discrimination was evaluated using Harrell's C statistic³³, a time-to-event area under the receiver operating characteristic curve (AUC ROC). ROCs graph the sensitivity versus specificity for the model against a particular outcome. (Fig. 1) The C statistic is a measure of goodness of fit, where 0 is no fit, 0.5 is chance ("flip of a coin") and 1 is perfect fit. 95% confidence intervals were calculated. Prediction calibration was assessed using the Brier score, which is a measure of the residual between predictions and observations. It is a measure of a model's prediction accuracy. Lower Brier scores indicate better calibration, where 0 is a perfect model, 0.25 is chance ("flip of a coin") and 1 is no predictive capability. Calibration plots were created using a binned risk method and plotted against the ideal prediction line where predicted is equal to observed (straight line, slope = 1). The models underestimate risk when the calibration plot lies above the ideal line (predicted probability less than observed) and overestimate risk when the calibration plot lies below the ideal line (observed probability less than predicted). Calibration plots included histograms of probability counts and grouped observations. Median and interquartile ranges were calculated for quantitative baseline demographics. All statistical analyses were performed using R version 4.0.4. TRIPOD reporting guidelines were followed, and 95% confidence intervals reported when possible.

Funding

Funding was provided by the National Institutes of Health (R01AR05517601A2) and the Scoliosis Research Society.

Results:

Two hundred sixty-eight ASLS patients enrolled, and 187 patients received operative treatment. Two-year follow up was completed by 169 (90%) operatively treated patients who were included in our analysis. Baseline demographics of included patients are summarized in Table 2. The ASLS-1 cohort was comprised primarily of Caucasian women undergoing primary lumbar scoliosis surgeries. There were no revision fusions and few gross sagittal plane deformities. ASLS-1 patients were older (ASLS: 60.1±9.0 yrs, ISSG-ESSG: 56.5±17.3, p=0.008) with larger coronal plane deformity (ASLS: 55.5±15.1°, ISSG-ESSG: 37.7±21.9°, p<0.001) and less sagittal plane deformity (ASLS C7SVA: 35.7±46.1, ISSG-ESSG: 59.4±71.1, p<0.001) than the ISSG-ESSG development set. ASLS-1 patients underwent longer surgeries (ASLS: 7.0±2.2 hours, ISSG-ESSG: 5.7±2.4, p<0.001) with more frequent interbody fusions (ASLS: 67%, ISSG-ESSG: 50.6%, p<0.001) and more frequent iliac fixation (ASLS: 91%, ISSG-ESSG: 54%, p<0.001). The ASLS-1 patients tended to have less pain/disability at baseline.

Adverse Events Prediction Models

Model performance statistics and observed rates of MC, RO, and RA are found in Table 3.

Major Complication

Model predictions for MC probabilities ranged from 13%–68% with a mean of 38.7% and a C statistic of 0.61 (0.52–0.69). The ROC is found in Fig. 1A. The observed rate of MC was 41.4%. The model overestimated the probability of MC in most cases, except for those with the lowest likelihood (approximately 20%) of sustaining a MC (Fig. 2A). The Brier score was 0.22. The histogram of probability distributions shows a concentration of patients with predicted probabilities of MC falling from 20–60%.

Repeat Operation

Model predictions for RO ranged from 9–54% with a mean of 30.1% and a C statistic of 0.59 (0.49–0.70). The observed rate of RO was 20.7%. The ROC curve for RO is found in Fig. 1B and calibration curve is shown in Fig. 2B. The model overestimated probabilities of RO and had a Brier score of 0.17. The histogram of probability distributions shows a concentration of patients with predicted probabilities of RO falling between 20–40%.

Readmission

Model predictions for RA ranged from 13–50% with a mean of 28.5%, overestimating the observed rate of 17.2%. ROC curve for RA is shown in Fig. 1C and the calibration curve is shown in Fig. 2C. RA had an C statistic of 0.56 (0.46–0.67) and a Brier score of 0.16. The histogram of probability distributions shows a concentration of patients with predicted probabilities of RA falling between 20–40%.

Discussion:

Prediction models as instruments for personalized informed decision-making may improve satisfaction and value-delivery in ASD surgeries but require validation across populations and the variety of ASD sub-types. We sought to examine the properties of abbreviated ASD prediction models for major complication, readmission, and revision surgery in a cohort of symptomatic lumbar scoliosis patients. Each model performed better than chance, with AUCs of approximately 0.6 and Brier scores under 0.25. The models overestimated the probabilities of readmission (observed = 20.7%, predicted mean = 30.1%) and repeat operation (observed 17.2%, predicted mean = 28.5%). Prediction of major complication was not as consistent as the models underestimated risk (observed 41.4%, predicted mean 38.7%) for the whole cohort. The calibration line (Figure 2A) crosses “perfect” and shows that the models underestimate risk for low-risk patients and overestimate risk for high-risk patients. We believe these data offer proof of concept, while emphasizing the need for refinement of these models prior to their adoption in clinical care.

The performances of the calculators were not grossly different from the development and validation performance previously reported from a more broadly inclusive ASD cohort.²⁷ In the ISSG-ESSG validation cohorts, C statistic (goodness of fit) estimates ranged from 0.5 to 0.76. Calibration (Brier score) was slightly better for the ISSG-ESSG cohort versus

the ASLS-1 cohort. This is likely due to the selection bias associated with the ASLS-1 cohort, which excluded patients with any prior spine surgery of larger magnitude than a one-level decompression. Unsupervised clustering of the ISSG-ESSG cohort identified three primary patient types: young primary, old primary, and old revision, where the major complication rate in the old primary group was approximately 40%.²⁸ The inclusion criteria of ASLS-1 tended to select “old primary” where the demographic data, surgical data (Table 2), complication rates and profiles for these patients are different from a multiply operated iatrogenic flatback patient. This poor calibration, due to selection bias, is seen by the tight distribution of the probability histograms and grouped observations found in the calibration plots where the breadth of probabilities and true events were not observed. The ASLS population does not contain patients at the tails of the probability distributions (i.e., those certain to have and not have MC/RO/RA). Models are better suited to making predictions when strongly predictive variables are heterogenous.³³ In the case of ASLS-1, there is homogeneity that affects the modeling, though this encourages the continued collection of data to refine the models.

Understanding patient preferences is paramount to shared decision making. Substantial work has discussed the importance of understanding a patient’s “preference phenotype” where their decision making is based upon both risk-tolerance and desired outcomes.³⁴ It is important to ensure that the experience of surgery and outcomes align with a patient’s expectations. Machine-learning algorithms will facilitate communication between the physician and patient to understand how the risks and expected outcomes of various treatment approaches align with their preference. Conjoint analysis, a method frequently used in marketing strategies, can be used to identify risk-benefit tradeoffs for patients. In conjoint analyses, patients are offered different options and expectations of outcomes and use these data to choose a treatment pathway.³⁵ The models created by the ISSG-ESSG offer personalized probability predictions to further optimize this advancement of the shared decision-making process.

Personalized prediction models are a natural progression given the advancement of machine-learning techniques and data availability. Data presented to patients is often in the form of a population average or some level of risk increase relative to patients with a particular attribute. For example, a patient with diabetes mellitus may have a 20% increased risk of deep wound infection, though this requires understanding the context of baseline risk and other confounding variables. While we believe these are good data and “talking points” with patients, we often fail to realize that “the median isn’t the message.”³⁶ The use of medians, means, and odds ratios fail to acknowledge the complexity of any single patient’s presenting problem or a surgeon’s individual experience with a procedure. Non-linear, machine-learning techniques allow us to combine patient and surgical data to offer the probabilities for outcomes such as major complication, readmission, and reoperation.

Prediction of patient-reported outcomes measures (PROM) is required if we are to optimize value in a shared decision-making process. Not only must we understand a patient’s dissatisfaction with their current state and tolerance for complication, but we must understand their desired PROM result. The ISSG-ESSG models have been used to predict responses to the components of the Scoliosis Research Society-22 questionnaire.²⁵

This will further refine the decision-making process to ensure the patient understands both the likelihood of complication and the likelihood of improving in domains important to them. This may increase the number of patients achieving the “minimum worthwhile effort” thereby improving value in adult spinal deformity surgery. Broad use of the minimum clinically important difference fails to acknowledge the risks and expectations of patients; a refined decision-making process might help determine those patients most likely to achieve the outcomes they desire.

This study is limited by the selection bias in the ASLS-1 cohort. By selecting “old primary” patients, we do not have a breadth of patient and surgery types to predict across the range of outcomes. The predictive properties of the models were modest at best in this validation study, with AUC closer to 0.5 than 1 and Brier scores closer to 0.25 than 0. However, in general, the models performed better than chance and we believe our results show proof-of-concept, encouraging further work in this field and refinement of the models. Selection bias also exists because surgeons operated on patients they deemed eligible for surgery and likely avoided those at the highest risk for poor outcomes. For example, active nicotine use and poorly controlled diabetes are often contraindications to complex, elective ASD reconstructions. As a result, there were few patients with these conditions in the development and validation cohorts. The omission of diabetes should not be interpreted to mean that blood sugar control is irrelevant, but rather it should be interpreted in the context of surgeons avoiding those patients felt to be the highest risk. It is unlikely that any model will ever include these patients. Prediction of outcomes in biological systems is complex, with bias and noise affecting estimates, but that a computer could perform better than a coin flip is evidence of potential. A “complete” counseling tool requires patient-reported outcomes measures, which we do not have here. Machine-learning models are often limited by a “black box” phenomenon where no simple regression equation can be offered for general use. We hope to make these calculators broadly available after further refinement. Finally, clinical studies of deployed models are required to show that patient satisfaction and value are improved. Given the results in other areas of musculoskeletal disease, we believe this will be the case in ASD as well.

Conclusions:

The abbreviated ISSG-ESSG prediction models performed better than chance for major complication, unplanned reoperation, and readmission after surgery for patients with adult symptomatic lumbar scoliosis. Calibration of the models was imperfect due to the patient cohort selected by the ASLS-1 study, which was comprised of mostly older patients undergoing primary fusion surgeries. These results emphasize the need for continued refinement and training of models if they are to be broadly applicable to ASD patients and reconstructions.

Acknowledgments

Sources of Support: National Institute of Arthritis and Musculoskeletal and Skin Diseases (Bridwell), Scoliosis Research Society (Bridwell), International Spine Study Group Foundation (Bess, Bridwell)

References

1. Passias PG, Jalai CM, Worley N, et al. Adult Spinal Deformity: National Trends in the Presentation, Treatment, and Perioperative Outcomes From 2003 to 2010. *Spine Deform.* Sep 2017;5(5):342–350. doi:10.1016/j.jspd.2017.02.002 [PubMed: 28882352]
2. Kebaish KM, Neubauer PR, Voros GD, Khoshnevisan MA, Skolasky RL. Scoliosis in adults aged forty years and older: prevalence and relationship to age, race, and gender. *Spine (Phila Pa 1976).* Apr 20 2011;36(9):731–6. doi:10.1097/BRS.0b013e3181e9f120 [PubMed: 20881515]
3. Schwab F, Dubey A, Gamez L, et al. Adult scoliosis: Prevalence, SF-36, and nutritional parameters in an elderly volunteer population. *Spine.* May 1 2005;30(9):1082–1085. doi:DOI 10.1097/01.brs.0000160842.43482.cd [PubMed: 15864163]
4. Kelly MP, Lurie JD, Yanik EL, et al. Operative Versus Nonoperative Treatment for Adult Symptomatic Lumbar Scoliosis. *The Journal of Bone and Joint Surgery.* Feb 20 2019;101(4):338–352. doi:10.2106/jbjs.18.00483 [PubMed: 30801373]
5. 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.* Spring 2014;14(1):67–77.
6. Youssef JA, Orndorff DO, Patty CA, et al. Current status of adult spinal deformity. *Global Spine J.* Mar 2013;3(1):51–62. doi:10.1055/s-0032-1326950 [PubMed: 24436852]
7. 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).* Sep 15 2009;34(20):2171–8. doi:10.1097/BRS.0b013e3181a8fdc8 [PubMed: 19752703]
8. Moal B, Lafage V, Smith JS, et al. Clinical Improvement Through Surgery for Adult Spinal Deformity: What Can Be Expected and Who Is Likely to Benefit Most? *Spine Deform.* Nov 2015;3(6):566–574. doi:10.1016/j.jspd.2015.04.004 [PubMed: 27927560]
9. Nuñez S, Pellisé F, Vila A, Domingo M, Sánchez Pérez-Grueso F, Acaroglu E. Risk factors and clinical impact of early unanticipated revision surgery in adult spinal deformity. *Eur Spine J.* 2015;24:770–771.
10. Yeramaneni S, Gum JL, Carreon LY, et al. Impact of Readmissions in Episodic Care of Adult Spinal Deformity: Event-Based Cost Analysis of 695 Consecutive Cases. *J Bone Joint Surg Am.* Mar 21 2018;100(6):487–495. doi:10.2106/jbjs.16.01589 [PubMed: 29557865]
11. Rihn JA, Currier BL, Phillips FM, Glassman SD, Albert TJ. Defining the value of spine care. *J Am Acad Orthop Surg.* Jul 2013;21(7):419–26. doi:10.5435/jaaos-21-07-419 [PubMed: 23818029]
12. Porter ME. What is value in health care? *N Engl J Med.* Dec 23 2010;363(26):2477–81. doi:10.1056/NEJMp1011024 [PubMed: 21142528]
13. Porter ME, Lee TH. From Volume to Value in Health Care: The Work Begins. *JAMA.* 2016;316(10):1047–1048. doi:10.1001/jama.2016.11698 [PubMed: 27623459]
14. Moor J. The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine.* 12/15 2006;27(4):87. doi:10.1609/aimag.v27i4.1911
15. Jayakumar P, Moore MLG, Bozic KJ. Value-based Healthcare: Can Artificial Intelligence Provide Value in Orthopaedic Surgery? *Clin Orthop Relat Res.* Aug 2019;477(8):1777–1780. doi:10.1097/corr.0000000000000873 [PubMed: 31335596]
16. Gum JL, Hostin R, Robinson C, et al. Impact of cost valuation on cost-effectiveness in adult spine deformity surgery. *Spine J.* Jan 2017;17(1):96–101. doi:10.1016/j.spinee.2016.08.020 [PubMed: 27523283]
17. Smith JS, Klineberg E, Lafage V, et al. Prospective multicenter assessment of perioperative and minimum 2-year postoperative complication rates associated with adult spinal deformity surgery. *J Neurosurg Spine.* Jul 2016;25(1):1–14. doi:10.3171/2015.11.Spine151036 [PubMed: 26918574]
18. Yagi M, Ames CP, Keefe M, et al. A cost-effectiveness comparisons of adult spinal deformity surgery in the United States and Japan. *Eur Spine J.* Mar 2018;27(3):678–684. doi:10.1007/s00586-017-5274-5 [PubMed: 28836012]

19. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med.* Sep 29 2016;375(13):1216–9. doi:10.1056/NEJMp1606181 [PubMed: 27682033]
20. Sellers MM, Merkow RP, Halverson A, et al. Validation of new readmission data in the American College of Surgeons National Surgical Quality Improvement Program. *J Am Coll Surg.* Mar 2013;216(3):420–7. doi:10.1016/j.jamcollsurg.2012.11.013 [PubMed: 23332220]
21. Khor S, Lavalley D, Cizik AM, et al. Development and Validation of a Prediction Model for Pain and Functional Outcomes After Lumbar Spine Surgery. *JAMA Surg.* Jul 1 2018;153(7):634–642. doi:10.1001/jamasurg.2018.0072 [PubMed: 29516096]
22. Lee MJ, Cizik AM, Hamilton D, Chapman JR. Predicting medical complications after spine surgery: a validated model using a prospective surgical registry. *Spine J.* Feb 1 2014;14(2):291–9. doi:10.1016/j.spinee.2013.10.043 [PubMed: 24239799]
23. Veeravagu A, Li A, Swinney C, et al. Predicting complication risk in spine surgery: a prospective analysis of a novel risk assessment tool. *J Neurosurg Spine.* Jul 2017;27(1):81–91. doi:10.3171/2016.12.Spine16969 [PubMed: 28430052]
24. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, et al. MySurgeryRisk: Development and Validation of a Machine-learning Risk Algorithm for Major Complications and Death After Surgery. *Ann Surg.* Apr 2019;269(4):652–662. doi:10.1097/sla.0000000000002706 [PubMed: 29489489]
25. Ames CP, Smith JS, Pellise F, et al. Development of predictive models for all individual questions of SRS-22R after adult spinal deformity surgery: a step toward individualized medicine. *Eur Spine J.* Sep 2019;28(9):1998–2011. doi:10.1007/s00586-019-06079-x [PubMed: 31325052]
26. Ames CP, Smith JS, Pellise F, et al. Development of Deployable Predictive Models for Minimal Clinically Important Difference Achievement Across the Commonly Used Health-related Quality of Life Instruments in Adult Spinal Deformity Surgery. *Spine (Phila Pa 1976).* Aug 15 2019;44(16):1144–1153. doi:10.1097/BRS.0000000000003031 [PubMed: 30896589]
27. Pellise F, Serra-Burriel M, Smith JS, et al. Development and validation of risk stratification models for adult spinal deformity surgery. *J Neurosurg Spine.* Jun 28 2019:1–13. doi:10.3171/2019.3.SPINE181452
28. Ames CP, Smith JS, Pellise F, et al. Artificial Intelligence Based Hierarchical Clustering of Patient Types and Intervention Categories in Adult Spinal Deformity Surgery: Towards a New Classification Scheme that Predicts Quality and Value. *Spine (Phila Pa 1976).* Jul 1 2019;44(13):915–926. doi:10.1097/BRS.0000000000002974 [PubMed: 31205167]
29. Brodke DS, Goz V, Lawrence BD, Spiker WR, Neese A, Hung M. Oswestry Disability Index: a psychometric analysis with 1,610 patients. *Spine J.* Mar 2017;17(3):321–327. doi:10.1016/j.spinee.2016.09.020 [PubMed: 27693732]
30. Bridwell KH, Cats-Baril W, Harrast J, et al. The validity of the SRS-22 instrument in an adult spinal deformity population compared with the Oswestry and SF-12: a study of response distribution, concurrent validity, internal consistency, and reliability. *Spine (Phila Pa 1976).* Feb 15 2005;30(4):455–61. doi:10.1097/01.brs.0000153393.82368.6b [PubMed: 15706344]
31. Jenkinson C, Coulter A, Wright L. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. *Bmj.* May 29 1993;306(6890):1437–40. doi:10.1136/bmj.306.6890.1437 [PubMed: 8518639]
32. Mannion AF, Bianchi G, Mariaux F, et al. Can the Charlson Comorbidity Index be used to predict the ASA grade in patients undergoing spine surgery? *Eur Spine J.* Sep 18 2020;doi:10.1007/s00586-020-06595-1
33. Pencina MJ, D'Agostino RB Sr., Evaluating Discrimination of Risk Prediction Models: The C Statistic. *JAMA.* 2015;314(10):1063–1064. doi:10.1001/jama.2015.11082 [PubMed: 26348755]
34. Fraenkel L, Nowell WB, Michel G, Wiedmeyer C. Preference phenotypes to facilitate shared decision-making in rheumatoid arthritis. *Ann Rheum Dis.* May 2018;77(5):678–683. doi:10.1136/annrheumdis-2017-212407 [PubMed: 29247126]
35. Shapiro LM, Eppler SL, Baker LC, Harris AS, Gardner MJ, Kamal RN. The Usability and Feasibility of Conjoint Analysis to Elicit Preferences for Distal Radius Fractures in Patients 55 Years and Older. *J Hand Surg Am.* Oct 2019;44(10):846–852. doi:10.1016/j.jhsa.2019.07.010 [PubMed: 31495523]

36. Kirkeboen G. “The median isn’t the message”: How to communicate the uncertainties of survival prognoses to cancer patients in a realistic and hopeful way. *Eur J Cancer Care (Engl)*. Jul 2019;28(4):e13056. doi:10.1111/ecc.13056

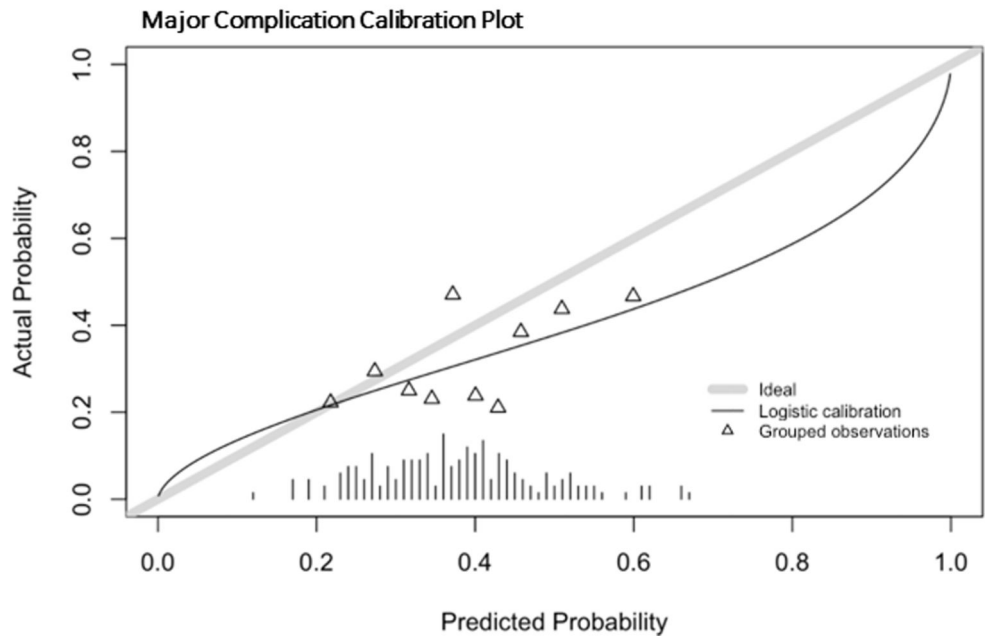
Author Manuscript

Author Manuscript

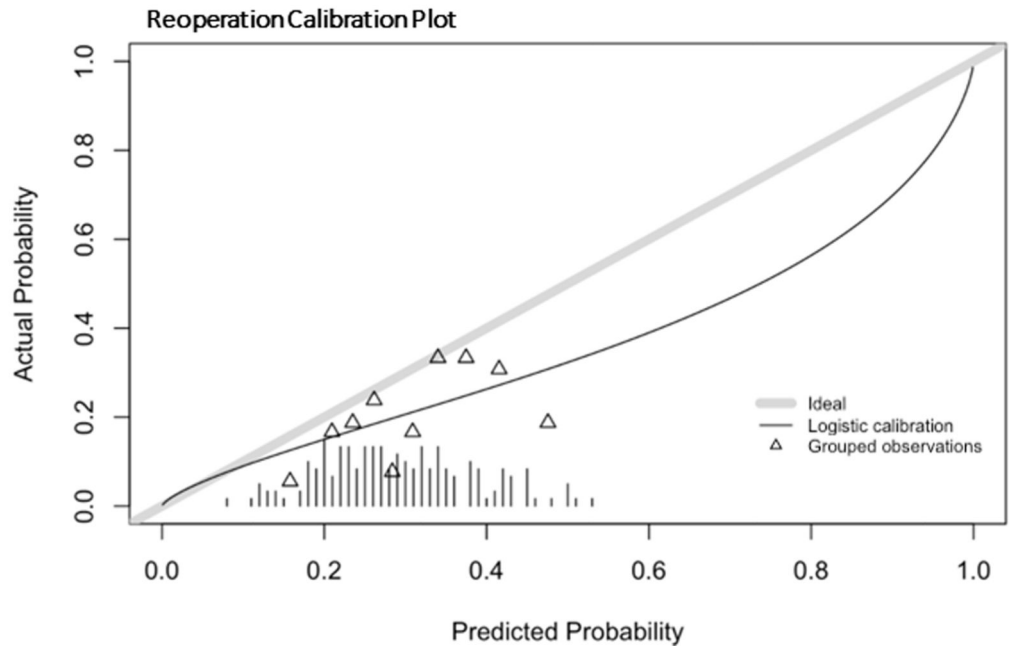
Author Manuscript

Author Manuscript

1A



1B



1C

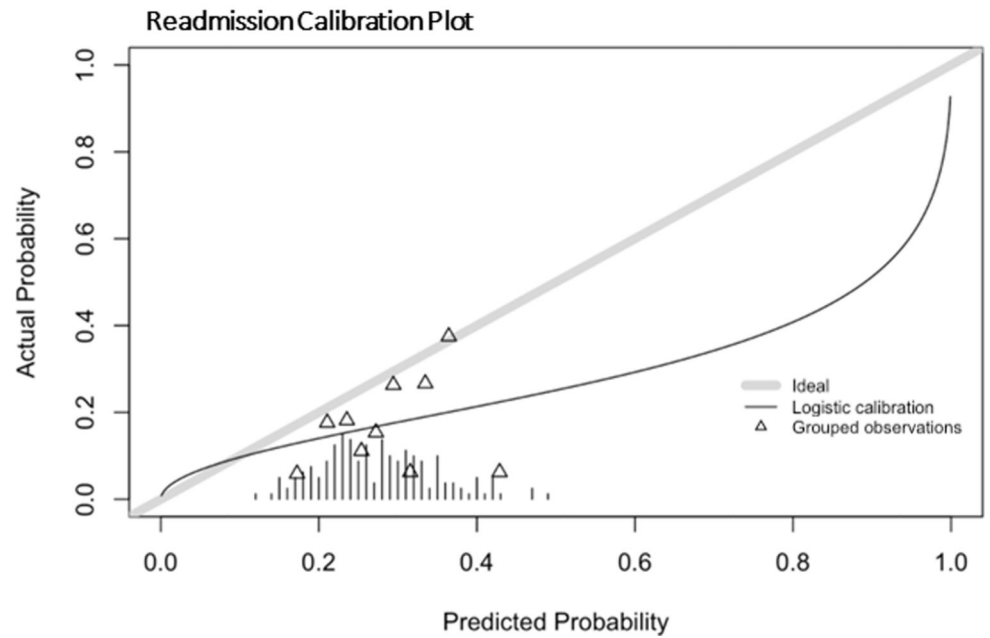
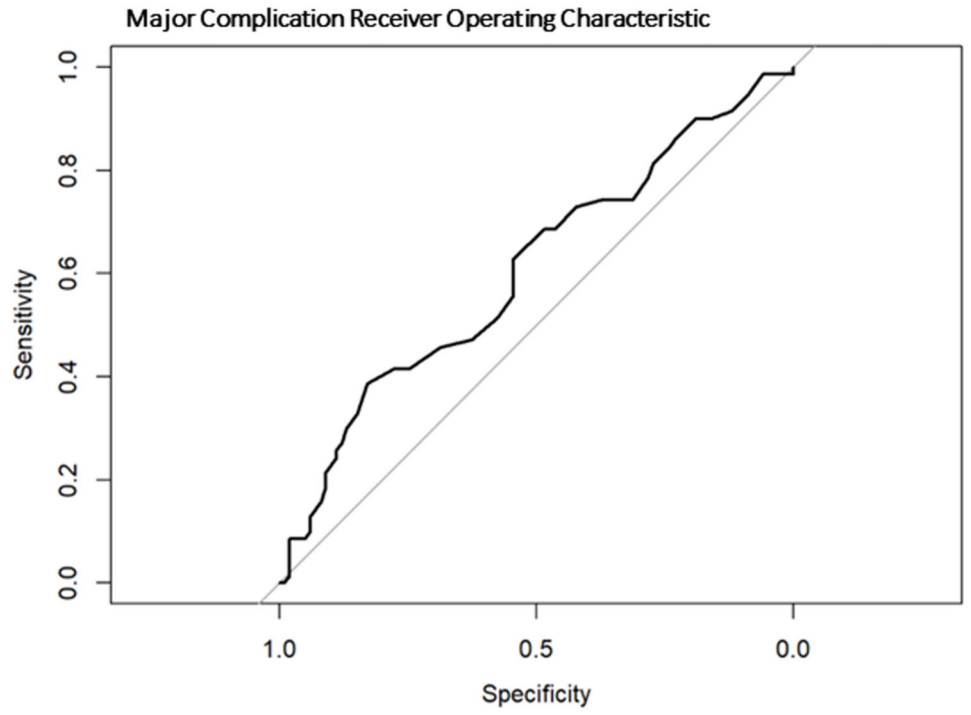
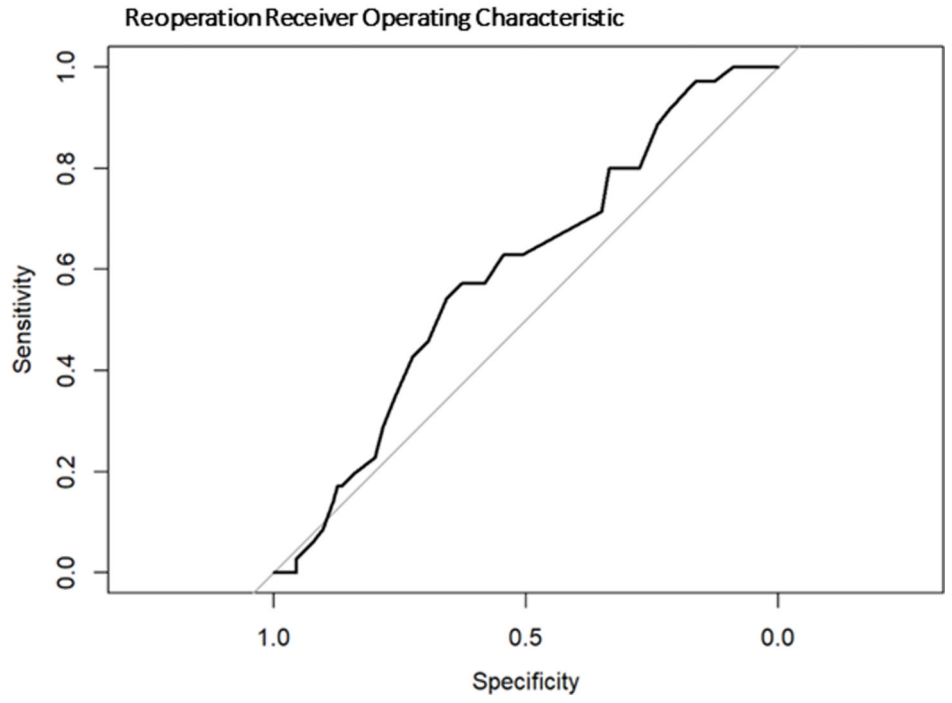


FIG. 1. Receiver operating characteristic curves for major complication (A), reintervention (B), unintended hospital readmission (C)

2A



2B



2C

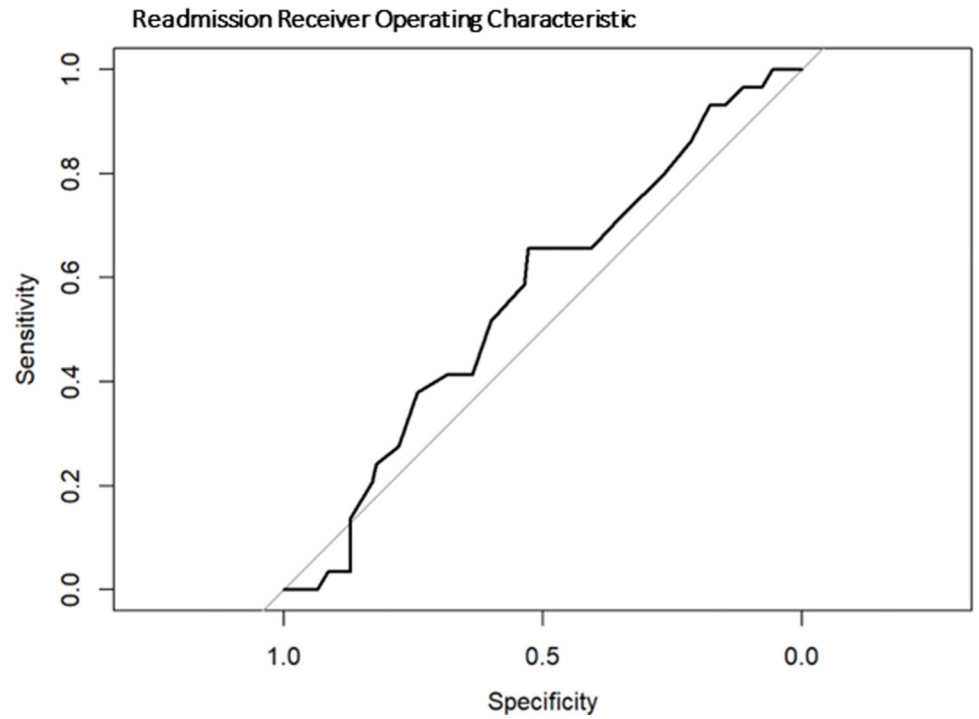


FIG. 2. Calibration curves for major complication (A), reintervention (B), unintended hospital readmission (C). Hatch marks form a histogram of events. Triangles represent grouped observation bins.

TABLE 1.

Patient factors utilized by predictive models

Demographic (12)	Radiographic (9) and HRQoL (4)	Surgical (14)
Age	Major curve location	Number of stages
Gender	Major curve Cobb angle	Antifibrinolytics
Prior spine surgery	Coronal balance	Levels fused
Height	Leg length discrepancy	Upper instrumented vertebra
Weight	L1-S1 angle	Lower instrumented vertebra
Smoking status	Pelvic tilt	Implant density (implants/level)
Heel walk	Sacral slope	Posterior rod material
Toe walk	Sagittal balance	Posterior rod diameter
Leg weakness	Global alignment	Multiple rods
ASA grade	ODI, walking	Posterior interbody fusions
NRS back pain	ODI, homemaking	Anterior interbody fusions
NRS leg pain	SRS-22r, medication usage	Levels decompressed
	SF-36v2, daily activities	Three column osteotomies
		Two column osteotomies

HRQoL = Health related quality of life; ASA= American Society of Anesthesiologists; NRS=Numeric Rating Scale; ODI= Oswestry Disability Index; SRS= Scoliosis Research Society; SF= Short Form

TABLE 2.

Baseline characteristics of ASLS and ISSG-ESSG Cohorts

Baseline Predictor	ASLS Cohort N=169	ISSG-ESSG Validation Cohort N=1289	p-value
Demographic Data			
Age (years)	60.1±9.0	56.5±17.3	0.008
Gender (Female)	151(89%)	1000(78%)	<0.001
Height (cm)	160.6±8.6	163±11.0	0.006
Weight (kg)	69.9±15.6	71.5±17.7	0.26
Any Prior Spine Surgery (Y)	19(11%)	544(42%)	<0.001
Major Coronal Cobb	55.5±15.1°	37.7±21.9°	<0.001
Pelvic Incidence	55.6±11.1°	55.5±13.0°	0.98
Pelvic Incidence-Lumbar Lordosis	19.0±18.3°	21.4±20.9°	0.16
Pelvic Tilt	24.2±8.9°	23.2±11.0°	0.26
C7 Sagittal Vertical Axis (mm)	35.7±46.1	59.4±71.1	<0.001
Surgical Data			
Estimated Blood Loss (mL)	2206±1676	1496±1315	<0.001
Surgical Time (hours)	7.0±2.2	5.7±2.4	<0.001
Number of Levels Fused	11.0±3.7	10.4±4.4	0.09
Number of Patients with Any Interbody Fusion	113(67%)	652(50.6%)	<0.001
Pelvic fixation performed (Y)	153(91%)	700(54%)	<0.001
Any Osteotomy Performed (Y)	106(63%)	769(59.7%)	0.44
Health Related Quality of Life			
Baseline Oswestry Disability Index	38.4±15.5	42.8±19.4	0.005
Baseline Mental Component Summary	49.6±11.5	43.8±12.9	<0.001
Baseline Physical Component Summary	33.2±9.8	33.4±9.9	0.80
SRS -22r Function	3.1±0.7	2.9±0.9	0.006
SRS-22r Pain	2.8±0.8	2.5±0.9	<0.001
SRS-22r Subscore	3.0±0.6	2.8±0.7	<0.001

SRS=Scoliosis Research Society

TABLE 3.

Discrimination and calibration statistics

Outcome	Predicted Probability (%)	Occurrence in ASLS (%)	Discrimination	Calibration		
			C Statistic	Brier Score	y-intercept	slope
Major Complication	38.7 (13, 68)	41.4	60.6% [52%, 69%]	21.8%	-0.50 (-0.95, -0.05)	0.62 (-0.09, 1.32)
Reoperation	30.1 (9, 54)	20.7	59.4% [49%, 70%]	17.3%	-0.74 (-1.51, 0.03)	0.72 (-0.13, 1.56)
Readmission	28.5 (13, 50)	17.2	56.4% [46%, 67%]	15.7%	-1.09 (-2.17, -0.02)	0.52 (-0.59, 1.63)

Adverse events are categorical and evaluated with c statistic. Predicted probability is reported as mean (range). Parentheses indicate 95% confidence intervals.