Supplementary Online Content

Khor S, Lavallee DC, Cizik AM, et al. Development and validation of a prediction model for pain and functional outcomes after lumbar spine surgery. *JAMA Surg*. Published online March 7, 2018. doi:10.1001/jamasurg.2018.0072.

eAppendix. Detailed description on the handling of missing data and the multiple imputation procedure and detailed information on our models' performance.

eFigure 1. Calibration plots of the development and Validation cohort.

eFigure 2. Receiver Operating Characteristic Curves for the development and validation cohorts.

eTable 1. Model coefficients

eTable 2. Calibration of the models

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix

1. Detailed description on the handling of missing data and the multiple imputation procedure

In order to determine how to handle the missing PROs data at 12 months, we first examined whether the data was missing at random (MAR), meaning that the likelihood of the outcome being missing depends only on the observed data. To do this, we performed logistic regressions on the missingness of ODI, NRS back pain, and NRS leg pain to examine the relationships between their missingness and other variables in our study. We found significant associations between some of the variables and the missingness of the follow-up PROs, suggesting that the follow-up data was likely to be MAR. Given this information, the multiple imputation approach was chosen over a complete case analysis where all incomplete data is removed from the analysis, since complete case analysis can cause bias of unknown size and direction and a reduction in power.[1,2] Although easier to implement, complete case analysis relies upon stronger missing data assumptions than multiple imputation modelling.

A multiple imputations procedure with 40 imputation iterations was performed using the Multiple Imputation by Chained Equations methods. In this iterative imputation approach each variable with missing data is modeled conditional upon the other variables in the data using a series of regression models.[3] Multiple predictions for each missing value are created, accounting for the statistical uncertainty in the imputations. The result is an augmented dataset containing 40 times the original sample size, all with complete data.

2. Detailed information on our models' performance

There is good to excellent agreement between the predicted probabilities and the observed in the development cohort. (Figure A1) The slope of the calibration plots were close to 1, with intercepts close to zero, indicating good calibration (Table A1). The biggest variation was seen in the model for ODI, where, when a line was fitted to the plot, the slope was 0.77, and the intercept was 0.13, indicating that the disability model tended to be more optimistic than what was actually observed, especially at low probabilities. There was no significant difference between the observed and model predictions, as suggested by the observed to expected ratios of 1.00, with confidence intervals spanning 1 for all three models. The AUCs ranged between 0.73 – 0.75 for ODI, NRS back and NRS leg pain improvement, reflecting good to excellent discrimination. (Figure A2)

The calibration in the validation cohort was good (Table A1). Observed-to-expected ratios were between 0.94 and 1.02, with confidence intervals spanning 1 for all three models, indicating no significant difference between the observed and model predictions. C-statistics was 0.66 for ODI improvement, 0.79 for back pain improvement, and 0.69 for the leg pain improvement model, demonstrating good discrimination. (Figure A2)

Figure 1: Calibration plots of the development and Validation cohort. Each dot reflects the predicted vs. observed probability for each group of 15 for the development and validation cohort. The 45 degrees line indicates perfect prediction.

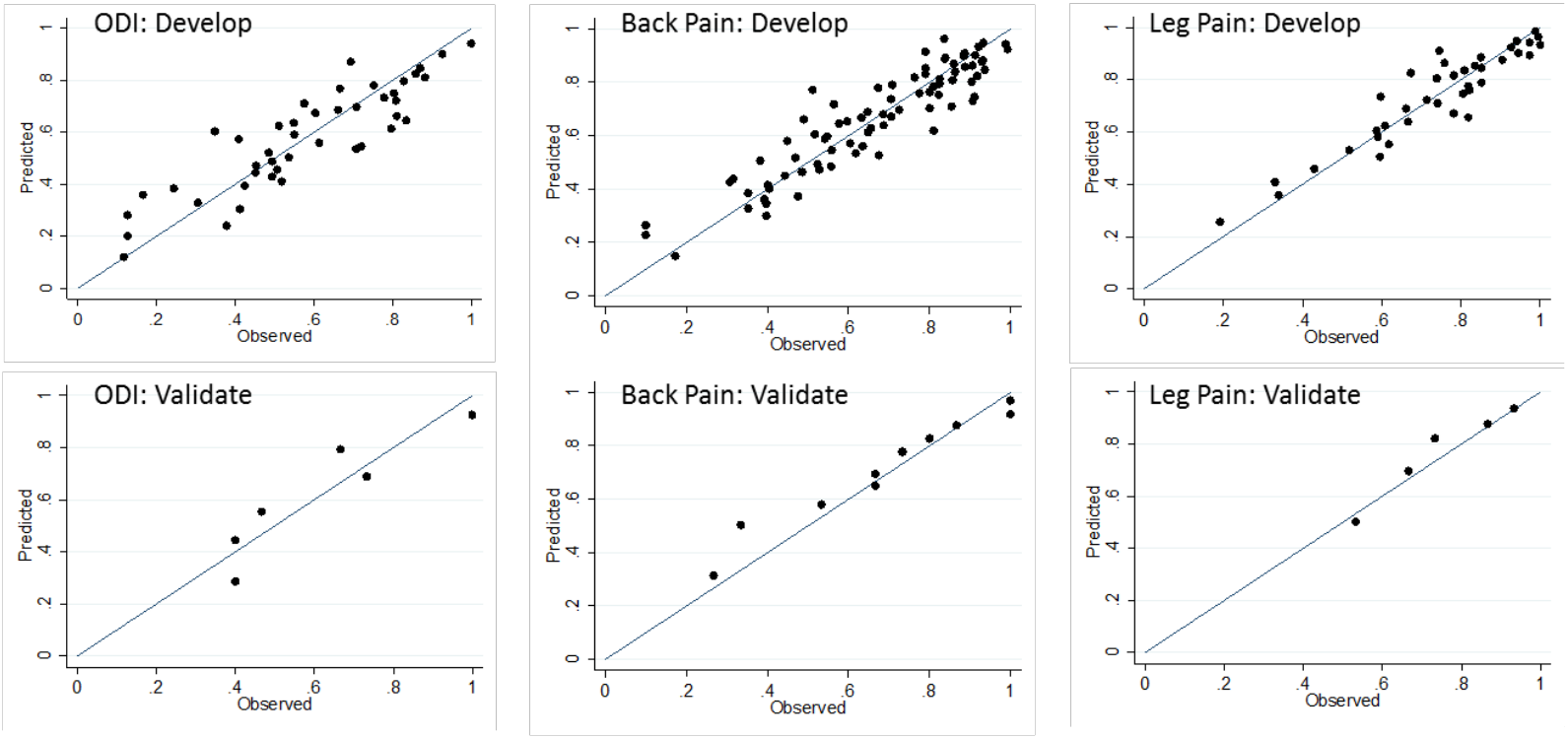

Figure 2: Receiver Operating Characteristic Curves for the development (green) and validation (blue) cohorts. The true positive rate is plotted against the false positive rate on each plot. The area under the ROC curve (AUC, or c-statistics) quantifies the predictive performance of each model. An area of 1 indicates perfect performance an area of 0.5 indicates no discriminative ability (shown on the plots by the 45 degrees line).

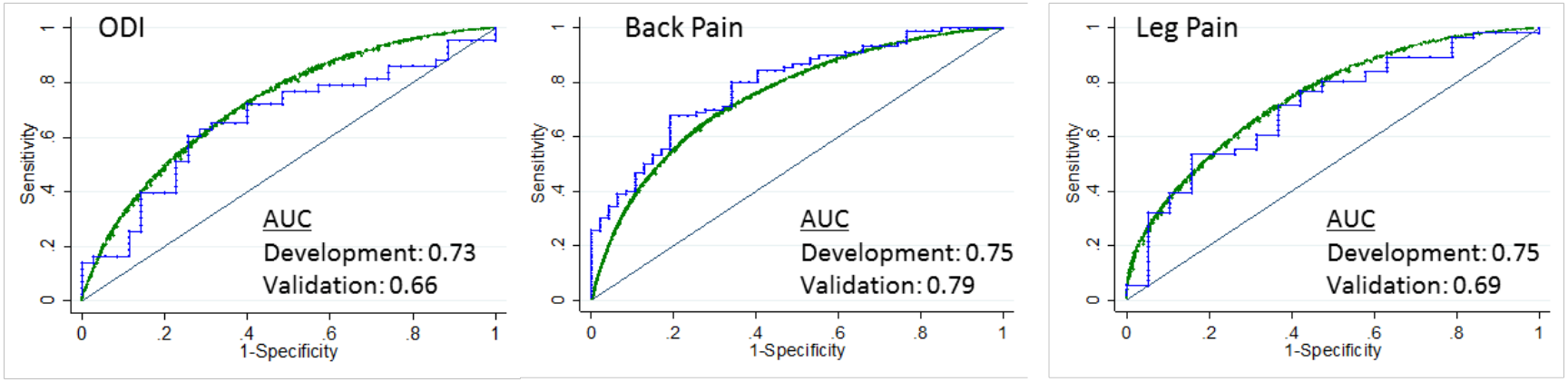

Table 1: Model coefficients. Probability of improvement in ODI, back NRS, and leg NRS can be computed by multiplying each patient's unique value for each parameter by the coefficient of the parameter, and then took the inverse logit of the total.

	ODI	NRS Back Pain	NRS leg pain
Age	-0.003	0.02	-0.01
Male	-0.08	-0.09	-0.22
Insurance (reference: Private)			
Medicaid	-0.96	-0.90	-0.29
Workers' Compensation	-1.61	-0.65	-0.73
Other	-0.30	-0.31	0.36
Race non-white (Reference: White)	-0.03	-0.12	-0.54
ASA 3 or higher	-0.17	-0.24	-0.42
Smoking Status (Reference: Never)			
Current	-0.85	-0.55	-0.45
Previous	-0.42	-0.21	-0.28
Prior Surgery	-0.50	-0.18	-0.02
Spondylolisthesis	0.56	0.49	0.26
Disc Herniation	0.50	0.11	0.48
Post Laminectomy/Failed Back Syndrome	-0.08	-0.06	-0.83
Stenosis	0.12	0.07	0.16
Pseudarthrosis	-1.06	-0.75	-0.52
Radiculopathy	-0.47	-0.03	-0.98
Prescription opiate use	0.05	-0.42	-0.32
Asthma	-0.61	-0.16	-0.14
Baseline ODI	0.05		
Baseline NRS back		0.43	
Baseline NRS leg			0.41
Constant	-0.89	-2.34	0.44

ASA, American Society of Anesthesiologists; ODI, Oswestry Disability Index; NRS, Numerical Rating Scale

Table 2: Calibration of the models

Cohort	Model	Calibration Slope	Calibration Intercept	Observed to Expected Ratio (95%CI)
Development	ODI	0.77	0.13	1.00 (0.93-1.07)
	NRS	0.84	0.11	1.00 (0.96-1.03)
	NRS leg	0.87	0.10	1.00 (0.96-1.03)
Validation	ODI	0.91	0.06	1.02 (0.8-1.24)
	NRS	0.80	0.16	0.94 (0.86-1.03)
	NRS leg	1.05	-0.02	0.98 (0.9-1.06)

CI, confidence interval; ODI, Oswestry Disability Index, NRS, Numerical Rating Scale