

Development and temporal validation of a prognostic model for one-year clinical outcome after decompression surgery for lumbar disc herniation

Lukas P Staub¹, Emin Aghayev², Veronika Skrivankova³, Sarah J Lord¹, Daniel Haschtmann², Anne F Mannion²

¹ The University of Sydney, NHMRC Clinical Trials Centre, Sydney, Australia

² Schulthess Klinik, Spine Center, Zurich, Switzerland

³ The University of Bern, Institute of Social and Preventive Medicine, Bern, Switzerland

Correspondence:

Lukas Staub, MD PhD

NHMRC Clinical Trials Centre, the University of Sydney

Locked bag 77, Camperdown NSW 1450

Australia

Lukas.staub@ctc.usyd.edu.au

Phone: +61 2 9562 5000

Acknowledgments

This research was supported by a research grant from the Mäxi Foundation, Switzerland. We thank all the patients of the Schulthess Klinik who have contributed data to the registry. We thank all present and past surgeons working at the Schulthess Klinik Spine Center for their compliance with the surgical forms of the Spine Tango Registry. We thank Dave O'Riordan, Gordana Balaban, Julian Amacker, Kirsten Clift, Sara Preziosa, Stéphanie Dosch, Riccardo Curatolo, and Selina Nauer for the administration of the surgical forms and patient-rated outcome measures in our registry and for the preparation of the data files for analysis. We thank Jana Neumann for her work on the statistical analysis plan. We thank Oliver Zumbrennen for the programming of the online tool.

Abstract

Purpose: Surgeons need tools to provide individualised estimates of surgical outcomes and the uncertainty surrounding these, to convey realistic expectations to the patient. This study developed and validated prognostic models for patients undergoing surgical treatment of lumbar disc herniation, to predict outcomes one year after surgery, and implemented these models in an online prediction tool.

Methods: Using the data of 1244 patients from a large spine unit, LASSO and linear regression models were fitted with 90% upper prediction limits, to predict scores on the Core Outcome Measures Index, and back and leg pain. Candidate predictors included sociodemographic factors, baseline symptoms, medical history, and surgeon characteristics. Temporal validation was conducted on 271 more recent patients at the same unit, by examining the proportion of observed outcomes exceeding the threshold of the 90% upper prediction limit (UPL), and by calculating mean bias and other calibration measures.

Results: Poorer outcome was predicted by obesity, previous spine surgery and having basic obligatory (rather than private) insurance. In the validation data, fewer than 12% of outcomes were above the 90% UPL. Calibration plots for the model validation showed values for mean bias <0.5 score points and regression slopes close to 1.

Conclusion: While the model accuracy was good overall, the prediction intervals indicated considerable predictive uncertainty on the individual level. Implementation studies will assess the clinical usefulness of the online tool. Updating the models with additional predictors may improve the accuracy and precision of outcome predictions.

Key words: lumbar disc herniation; decompression, surgical; patient reported outcome measures; prediction model

Introduction

Patients are more likely to be satisfied with their lumbar disc herniation surgery if their pre-operative expectations for pain and functional improvement are met.[1,2] Thus, surgeons need tools to provide individualised estimates of patient-oriented outcome, and the uncertainty surrounding such estimates, to establish realistic patient expectations.

Various baseline patient characteristics have been associated with surgical outcomes. A review of studies arising from the Spine Patient Outcomes Research Trial (SPORT) found that, in patients undergoing surgery for herniated disc, a positive surgical outcome was associated with preoperative leg pain (as opposed to back pain), absence of sensory deficit, posterolateral herniation, sequestered/extruded disc and not having received prior physical therapy.[3] A systematic review confirmed these findings and identified additional predictors of a negative outcome, including intact annulus fibrosus, long duration of sick leave, worker's compensation and high levels of preoperative symptoms.[4]

To apply this evidence in clinical practice, there is growing interest in the development of mathematical models that combine multiple prognostic factors for prediction of patient outcome, and incorporate these into computerised prognostic tools. The methods for the development [5,6], validation [7], and application [8] of such prognostic models are well established. Numerous models have been developed for the prognosis of low back pain in primary care [9] but few exist in relation to spine surgical treatment in tertiary care settings. Vroomen et al [10] developed a model to predict which patients initially presenting with clinical findings of nerve root compression will eventually undergo lumbar disc surgery. Two recent studies presented preoperative nomograms to predict patient-specific clinical and quality of life outcomes in patients undergoing cervical spine surgery [11] and a prediction model for pain and functional outcomes after lumbar spinal surgery [12]. However, additional models for more homogeneous diagnostic patient groups are needed to allow better targeted outcome predictions in clinical practice.

The aim of this study was to develop prognostic models for patients undergoing surgical treatment of lumbar disc herniation to predict scores on a multidimensional outcome measure and back and leg pain scores, one year after surgery. Model performance was then assessed using temporal validation. Based on these prognostic models, an online tool was developed for use in clinical practice.

Methods

Data source

This study was conducted following the TRIPOD Statement for transparent reporting of multivariable prediction models for individual prognosis or diagnosis.[13]

The study was carried out by the Schulthess Klinik in Zürich, a large tertiary spine unit in Switzerland, using data from our local spine surgery outcomes database nested within the framework of the EUROSPINE Spine Tango Spine Surgery Registry (<https://www.eurospine.org/spine-tango.htm>). Data had been collected prospectively, using the Spine Tango Surgery forms (2005, 2006 and 2011 versions) and patient-based Core Outcome Measures Index (COMI) forms, from consecutive patients undergoing surgery by specialised neurosurgeons and orthopaedic spine surgeons in our spine centre, part of a tertiary care orthopaedic hospital.

The COMI is a validated questionnaire [14,15] enquiring about pain, function, symptom-specific well-being, quality of life, and disability in relation to the back problem. It is recommended to promote the standardisation of outcome measurements in surgical studies and registries.[16]

Approval for this study was obtained from the local ethics committee (reference number KEK-ZH-2014-0418). The data supporting this study are not publicly available. However, applicants may obtain Ethical approval from EUROSPINE to use registry data for specific research projects.

Participants

All patients who had undergone elective surgery for lumbar disc herniation, as defined by the official algorithm for diagnostic groups developed by EUROSPINE (http://www.eurospine.org/cm_data/def_of_degen_patho.pdf) between January 2005 and December 2016, and had completed COMI forms pre-operatively and one year post-operatively, were eligible for inclusion. Patients were excluded if they had unknown morbidity state (ASA), body mass index or smoking status at baseline; or if no pre or post-operative patient questionnaire was available (lost to follow-up).

The development dataset comprised patients operated on between 17th January 2005 and 30th January 2015. The validation dataset comprised patients operated on between 4th February 2015 and 23rd December 2016.

Outcome

Outcomes were COMI score, leg pain, and back pain recorded 12 months after surgery, each scored from 0 to 10. The predictions were calculated on the same scale.

Predictors

Fifteen candidate predictors were available in the documented data for the development of the prediction models. They included sociodemographic factors, baseline symptoms, medical history, and surgeon characteristics (**Table 1**).

The Spine Tango Surgery forms were used to document patient age and sex, morbidity state (ASA physical status), length of preoperative conservative treatment.[17] Patient insurance class (three classes in Switzerland: private; semi-private; basic obligatory), body mass index (BMI), smoking status, and surgeon characteristics were extracted from the hospital information system.

Validation

The predictor and outcome variables were identical in the development and validation data. In accordance with the TRIPOD guidelines [13], the performance of the models was evaluated with the validation data. We used the validation dataset to make individual outcome predictions (using the models obtained with the development dataset) and compared the predictions with the observed outcomes.

Statistical analysis

Using the development dataset, we built three prediction models using the LASSO (least absolute shrinkage and selection operator) method [18] and linear regression. As some of the collected variables were expected to be highly collinear, we used the LASSO to select variables most relevant for predicting the individual outcomes. Since prediction intervals are not readily available for the LASSO models, we then fitted the variables selected by LASSO with linear regression models in order to obtain new predictions and prediction limits. For more details on LASSO regression, see Online Resource 1.

For the validation set, we applied the same inclusion criteria to patients operated on more recently in the same clinic. For each of the three modelled outcomes, we calculated predicted values for all individuals in the validation dataset and computed 90% upper prediction limits, representing the score value that would be exceeded by 1 in 10 patients with the same predicted score.

To assess our models, we first examined the proportion of observed outcomes exceeding the threshold of the 90% upper prediction limit (UPL). If the threshold was accurate, we would expect around 10% of observed outcomes from the validation set to exceed this limit.

We also calculated a variety of measures commonly used for the evaluation of prediction models by comparing predicted and observed outcomes, including Mean Bias (MB), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Pearson correlation coefficient (r) and R-squared. The latter represents how much of the outcome variation is explained by the prediction model, i.e. it assesses the predictive capacity of the model. We further calculated the Index of Agreement (IA), introduced by Willmott [19], and Factor 2. The IA ranges between 0 and 1 and decreases when the predicted and observed outcomes are inconsistent with respect to the mean outcome (i.e. if they lie on different sides of the mean outcome). Factor 2 (FACT2) is the proportion of predicted values that differ from the observed values by less than a factor of 2.

All statistical analyses were performed in R version 3.5.2, using the basic package and glmnet.

Online tool

To present the prediction models in an accessible format for surgeons and patients, we developed a publicly accessible online tool that displays the estimated outcomes given the individual characteristics of a patient. The 90% upper prediction limits for the predicted COMI and pain scores were chosen to permit a comprehensible interpretation of the predictions and range of uncertainty.

The tool was implemented as a website in the style of a single-page application, which allows for simple dynamic rewriting of the page content without the need for server interaction. The application was programmed in JavaScript, using open source libraries JQuery and D3 Data-Driven Documents (D3.js).

Results

Participants

Of 1796 patients undergoing surgery for lumbar disc herniation, 1244 (69%) patients had all predictor and outcome information available and were included for model development (**Figure 1**).

Forty-three percent of patients were female, and the overall mean age was 50 years (range 15 – 88 years) (**Table 1**). More than half (54%) had basic obligatory health insurance, 55% were generally healthy (ASA 1), the majority were of normal weight (BMI 20 – 25 kg/m²), and 31% were smokers.

The chief symptom (reason for surgery) declared on the patient questionnaire was leg pain in 55% of patients.

Model development and specification

In the development dataset, the mean preoperative scores were 7.7 (SD 1.7) for COMI, 6.6 (SD 2.6) for leg pain, and 4.5 (SD 2.9) for back pain. Postoperative, the average scores had decreased to 3.2 (SD 2.8) for COMI, 2.4 (SD 2.7) for leg pain, and 2.7 (SD 2.6) for back pain. The predictions obtained with the Lasso regression and the corresponding linear regression models were nearly identical, supporting our approach of using the linear regression models for the subsequent analyses (Online Resource 1).

Figure 2 shows the ordered model-based predictions for the three (COMI, back pain, leg pain) linear regression models, along with the 80%, 90% and 95% upper prediction limits. In all three models, the 90% upper prediction levels were generally approximately 3 score points higher than the predictions.

Table 2 provides the non-standardised linear regression coefficients for the COMI, back pain and leg pain prediction models. These determine the equations for the outcome predictions. The model coefficients represent the size of effect per unit change in each predictor, within a given model. Being obese (BMI 31-35 kg/m²) and having had more than one previous spine surgery were both strong predictors of a poorer outcome (i.e. were associated with a higher score for the predicted outcome) in all models. Conversely, private and semi-private insurance status, compared with basic obligatory insurance, were both associated with a better outcome (i.e. were associated with a lower score for the predicted outcome). Patient sex did not appear to predict outcome in any of the models. Beyond these variables, different predictors were retained for the three prediction models.

Online prediction tool

The model is publicly available as an online tool (Online Resource 2), which can be accessed here: <https://linkup.kws.ch/prognostic>.

Temporal validation of models

In the validation dataset (n=364), the fitted lines of predicted versus observed predictions generally lay slightly below the diagonal line that represented perfect agreement (**Figure 3**). The slopes of the fitted lines were 0.96 in the COMI model, 0.91 in the back pain and 0.75 in the leg pain models. The proportions of outcomes above the 90% UPL were 10.0%, 11.4% and 11.8% for the COMI score, back pain and leg pain models (**Figure 3**).

The measures for model validation (**Table 3**) confirm the validation plots. For all three models we observed a negative bias. The model for back pain had the smallest Mean Bias (-0.14), followed by that for the COMI score (-0.41) and leg pain (-0.44). The Mean Absolute Error and Root Mean Squared Error were also smallest for the back pain model, and its R-squared was the highest (0.19), compared with the models for COMI score (R-sq=0.17) and leg pain (R-sq=0.06). The Pearson correlation coefficient was similar for COMI score (r=0.44) and back pain (r=0.43) models, and lower for the leg pain model (r=0.31). The Index of Agreement was also higher for the COMI score and back pain models (0.58) than for the leg pain model (0.48). The proportions of predicted values that differed from the observed values by less than a factor of 2 (FACT2) were around 50% for COMI and back pain models and 38% for the leg pain model.

Overall, the model predicting back pain was the best performing model, with the smallest values for MB, MAE and RMSE, and the highest R-squared.

Discussion

We present the development and validation of a prognostic online tool for predicting one-year clinical outcome after decompression surgery for lumbar disc herniation. We propose the use of the modelled predictions to inform patient expectations based on their own particular condition and circumstances, in the hope of improving the patient experience and satisfaction with surgery.

The achieved Mean Bias of less than 0.5 score points in the model validation is quite small, indicating that the predicted scores were on average close to the actual outcomes. **Figure 3** shows that the models are also reasonably well calibrated, as the linear regression slopes between predicted and observed values are close to 1. The accuracy of our models is good overall. However, given the large prediction intervals, the predictive capacity of the models is rather low. This means that although the actual individual patient outcomes are *on average* as predicted and the 90% UPL is correct, the predictive uncertainty is quite large. The R-squared values of less than 0.2 indicate that the models only explain a small percentage of the variability of the predicted outcomes around their mean. While low R-squared values are not uncommon in medical regression models [20], this suggests that some important factors predicting surgical outcomes were not included in our models. In particular, more information is needed on psychological factors, which are consistent predictors of surgical outcome [21]. A new patient-rated tool is being developed for the Spine Tango registry, which includes single items for each of the domains depression, anxiety, catastrophizing and fear avoidance beliefs.[22] Behavioural, socioeconomic, occupational and societal flags may also be

worthy of further investigation. In addition, information on clinical examination (e.g. Lasegue sign [23]) and imaging features [24] could potentially improve the goodness-of-fit of the models.

Still, the models provide the best available estimation of the likely outcomes for individual patients with herniated disc undergoing surgery in our hospital, based on the information routinely documented for the Spine Tango Registry. Qualitative research has shown that the satisfaction of patients with spine surgery depends on the level of pain and daily function achieved compared with preoperative expectations.[1] Patients are also often overly optimistic regarding the surgical outcome.[25] If these expectations are not met, patients are unlikely to be satisfied with the result. Hence, the wide prediction intervals displayed in the online tool may be an important input when setting realistic expectations of the surgical outcome and goal setting for rehabilitation.

The primary limitation of this observational study is that we know nothing about likely treatment outcomes in a comparator group receiving conservative treatment. The online tool can therefore not serve to inform clinical decisions for treatment selection. However, if it can be assumed that the symptoms are unlikely to resolve or improve without surgical treatment, and an indication for surgery has been established, then the tool should provide important prognostic information regarding the likely outcome of surgery, including a best and worst case scenario at the prediction limits. Another limitation is that the models were developed and validated for only one high-volume spine unit in the Swiss health care system. The generalisability of the prediction tool cannot be guaranteed, as patients in different settings may have largely different results. Future research should establish whether the prediction models presented here are useful in clinical practice. Implementation studies measuring whether and how the use of the online tool changes practice can give us this information.[8] If deemed useful, external validation studies in other clinics, using the model specifications of our study, will be needed. Such studies could use the pooled data from the EUROSPINE Spine Tango registry. Recalibration of some of the model coefficients may be necessary in order to improve the accuracy of predictions in other clinics. For application in an international setting, some predictors may be less relevant or not available (e.g. insurance type). The present models could be updated with new variables for these situations.

Even if the models presented here can be shown to be useful in practice, more work is needed to identify and collect information on additional predictors of surgical benefit. These factors can be used to update the models, further improving the accuracy and precision of outcome predictions.

References

1. Archer K, Bird ML, Haug C, Coronado R, Wegener S, Devin CJ, Schlundt D (2015) Patients' Experience and Expectations of Lumbar Spine Surgery for Degenerative Conditions: A Qualitative Study. *The Spine Journal* 15 (10, Supplement):S99-S100. doi:10.1016/j.spinee.2015.07.046
2. Ronnberg K, Lind B, Zoega B, Halldin K, Gellerstedt M, Brisby H (2007) Patients' satisfaction with provided care/information and expectations on clinical outcome after lumbar disc herniation surgery. *Spine* 32 (2):256-261. doi:10.1097/01.brs.0000251876.98496.52
3. Koerner JD, Glaser J, Radcliff K (2015) Which Variables Are Associated With Patient-reported Outcomes After Discectomy? Review of SPORT Disc Herniation Studies. *Clin Orthop Relat Res* 473 (6):2000-2006. doi:10.1007/s11999-014-3671-1
4. Wilson CA, Roffey DM, Chow D, Alkherayf F, Wai EK (2016) A systematic review of preoperative predictors for postoperative clinical outcomes following lumbar discectomy. *Spine J* 16 (11):1413-1422. doi:10.1016/j.spinee.2016.08.003
5. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG (2009) Prognosis and prognostic research: what, why, and how? *BMJ* 338:b375. doi:10.1136/bmj.b375
6. Royston P, Moons KG, Altman DG, Vergouwe Y (2009) Prognosis and prognostic research: Developing a prognostic model. *BMJ* 338:b604. doi:10.1136/bmj.b604
7. Altman DG, Vergouwe Y, Royston P, Moons KG (2009) Prognosis and prognostic research: validating a prognostic model. *BMJ* 338:b605. doi:10.1136/bmj.b605
8. Moons KG, Altman DG, Vergouwe Y, Royston P (2009) Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 338:b606. doi:10.1136/bmj.b606
9. Haskins R, Osmotherly PG, Rivett DA (2015) Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review. *J Clin Epidemiol* 68 (7):821-832. doi:10.1016/j.jclinepi.2015.02.003
10. Vroomen PC, de Krom MC, Knottnerus JA (2000) When does the patient with a disc herniation undergo lumbosacral discectomy? *J Neurol Neurosurg Psychiatry* 68 (1):75-79. doi:10.1136/jnnp.68.1.75
11. Lubelski D, Alentado V, Nowacki AS, Shriver M, Abdullah KG, Steinmetz MP, Benzel EC, Mroz TE (2018) Preoperative Nomograms Predict Patient-Specific Cervical Spine Surgery Clinical and Quality of Life Outcomes. *Neurosurgery* 83 (1):104-113. doi:10.1093/neuros/nyx343
12. Khor S, Lavalley D, Cizik AM, Bellabarba C, Chapman JR, Howe CR, Lu D, Mohit AA, Oskouian RJ, Roh JR, Shonnard N, Dagal A, Flum DR (2018) Development and Validation of a Prediction Model for Pain and Functional Outcomes After Lumbar Spine Surgery. *JAMA Surg* 153 (7):634-642. doi:10.1001/jamasurg.2018.0072
13. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 162 (1):55-63. doi:10.7326/M14-0697
14. Ferrer M, Pellise F, Escudero O, Alvarez L, Pont A, Alonso J, Deyo R (2006) Validation of a minimum outcome core set in the evaluation of patients with back pain. *Spine* 31 (12):1372-1379; discussion 1380. doi:10.1097/01.brs.0000218477.53318.bc

15. Mannion AF, Porchet F, Kleinstuck FS, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective. Part 1: the Core Outcome Measures Index in clinical practice. *Eur Spine J* 18 Suppl 3:367-373. doi:10.1007/s00586-009-0942-8
16. Mannion AF, Elfering A, Staerke R, Junge A, Grob D, Semmer NK, Jacobshagen N, Dvorak J, Boos N (2005) Outcome assessment in low back pain: how low can you go? *Eur Spine J* 14 (10):1014-1026. doi:10.1007/s00586-005-0911-9
17. Roder C, Chavanne A, Mannion AF, Grob D, Aebi M (2005) SSE Spine Tango--content, workflow, set-up. www.eurospine.org-Spine Tango. *Eur Spine J* 14 (10):920-924. doi:10.1007/s00586-005-1023-2
18. Tibshirani R (1996) Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met* 58 (1):267-288
19. Willmott CJ (1981) On the Validation of Models. *Physical Geography* 2 (2):184-194. doi:10.1080/02723646.1981.10642213
20. Hamilton DF, Ghert M, Simpson AH (2015) Interpreting regression models in clinical outcome studies. *Bone Joint Res* 4 (9):152-153. doi:10.1302/2046-3758.49.2000571
21. Mannion AF, Elfering A (2006) Predictors of surgical outcome and their assessment. *Eur Spine J* 15 Suppl 1:S93-108. doi:10.1007/s00586-005-1045-9
22. Mannion AF, Mariaux F, Fekete TF, Haschtmann D, Loibl M, Becker H-J, Kleinstuck FS, Porchet F, Jeszenszky D, Elfering A (2018) Time for one more question? A simple "yellow flag" tool to assess psychological factors in patients undergoing spine surgery. *Eur Spine J* (27 (Suppl 5)):670
23. Kohlboeck G, Greimel KV, Piotrowski WP, Leibetseder M, Krombholz-Reindl M, Neuhofer R, Schmid A, Klinger R (2004) Prognosis of multifactorial outcome in lumbar discectomy: a prospective longitudinal study investigating patients with disc prolapse. *Clin J Pain* 20 (6):455-461
24. Carragee EJ, Han MY, Suen PW, Kim D (2003) Clinical outcomes after lumbar discectomy for sciatica: the effects of fragment type and anular competence. *J Bone Joint Surg Am* 85 (1):102-108
25. Mannion AF, Junge A, Elfering A, Dvorak J, Porchet F, Grob D (2009) Great expectations: really the novel predictor of outcome after spinal surgery? *Spine* 34 (15):1590-1599. doi:10.1097/BRS.0b013e31819fcd52

Figure legends

Figure 1 Study profile

Figure 2 Modelled predictions for COMI score (a), back pain (b) and leg pain (c). Individual predictions are ordered (black line) with corresponding 80%, 90% and 95% upper prediction limits

Figure 3 Temporal validation of prediction models for COMI score (a), back pain (b) and leg pain (c)

Table 1 Potential predictors and their distribution in the sample (n=1 244)

Predictor Categories or Units	Distribution N (%) or Mean (SD)
Sex (female)	537 (43%)
Age (years)	50 (SD 14)
Extent of lesion: 1 level (vs >1 level)	1014 (82%)
Previous conservative treatment	
no previous treatment	155 (14%)
<6 months	565 (50%)
6-12 months	176 (16%)
>12 months	231 (21%)
Previous surgeries: none (vs ≥1)	937 (75%)
Surgeon seniority status	
Senior	511 (41%)
Intermediate	521 (42%)
Junior	212 (17%)
Surgeon specialisation	
neuro surgeon	916 (74%)
orthopaedic surgeon	328 (26%)
Which of the following problems troubles you most?	
back pain	232 (19%)
leg pain	681 (55%)
sensory disturbances	325 (26%)
other	6 (0%)
Insurance type	
private	270 (22%)
semi-private	293 (23%)
basic obligatory	681 (55%)
Morbidity state (ASA)	
no disturbance	696 (56%)
mild/moderate	456 (37%)
severe	92 (7%)
BMI (kg/m ²)	
<20	82 (7%)
20-25	606 (49%)
26-30	419 (34%)
31-35	94 (7%)
>35	43 (3%)
Current smoker	392 (32%)
Preop COMI score	7.7 (SD 1.7)
Preop back pain	4.5 (SD 2.9)
Preop leg pain	6.6 (SD 2.6)

Table 2 Linear regression coefficients (non-standardised)

Predictor	Reference	Calculation	Coefficients		
			COMI	Back pain	Leg pain
Model intercept			2.037	1.620	1.223
Sex	Male	Female	0	0	0
Age	50 years	Per 10 years	0	-0.048	0.038
Extent of lesion	1 segment	2 segments	0.234	0.272	0
Previous surgeries	None	One	0.549	0.225	0.517
		More than one	0.877	0.590	0.921
Surgeon seniority status	Senior consultant	Intermediate (fellow)	0	0.190	0.034
		Registrar	0	0.232	0
Surgeon specialisation	Orthopaedic surgeon	Neuro surgeon	0	0	0.197
Chief complaint	Leg pain	Back pain	0.393	0.573	0
		Sensory disturbance	0.489	0.515	0.433
		Other	0	0.838	0.696
Insurance	Basic obligatory	Private	-0.746	-0.491	-0.513
		Semi-private	-0.610	-0.503	-0.231
Morbidity state	No disturbance	Mild/moderate	0.333	0.250	0.259
		Severe morbidity	0.452	0	0.360
BMI	20-25	<20	0.288	0.311	0
		26-30	0.488	0.408	0.342
		31-35	1.130	0.901	0.798
Current smoker	No	Yes	0.299	0.116	0.176
Preop COMI score	5	Per 1 score point	0.187	0.119	0.093
Preop back pain	5	Per 1 score point	0.185	0.234	0.182
Preop leg pain	5	Per 1 score point	0.072	0.086	0.155
Previous treatment	No Treatment	Per each (initiated) 6-month period	0	0.090	0.119

Table 3 Measures of temporal model validation

Validation measure	Model		
	COMI score	Back pain	Leg pain
Mean Bias (MB)	-0.41	-0.14	-0.44
Mean Absolute Error (MAE)	2.04	1.83	2.03
Root Mean Squared Error (RMSE)	2.42	2.25	2.44
R-squared	0.17	0.19	0.06
Pearson correlation coefficient (r)	0.44	0.43	0.31
Index of Agreement (IA) ^a	0.58	0.58	0.48
Factor 2 (FACT2) ^b	0.49	0.51	0.38

^a) IA ranges between 0 and 1 and decreases when the predicted and observed outcomes are inconsistent with respect to the mean outcome, i.e. if they lie on different sides of the mean outcome.

^b) FACT2 is the proportion of predicted values that differ from the observed values by less than a factor of 2.

Figure 1

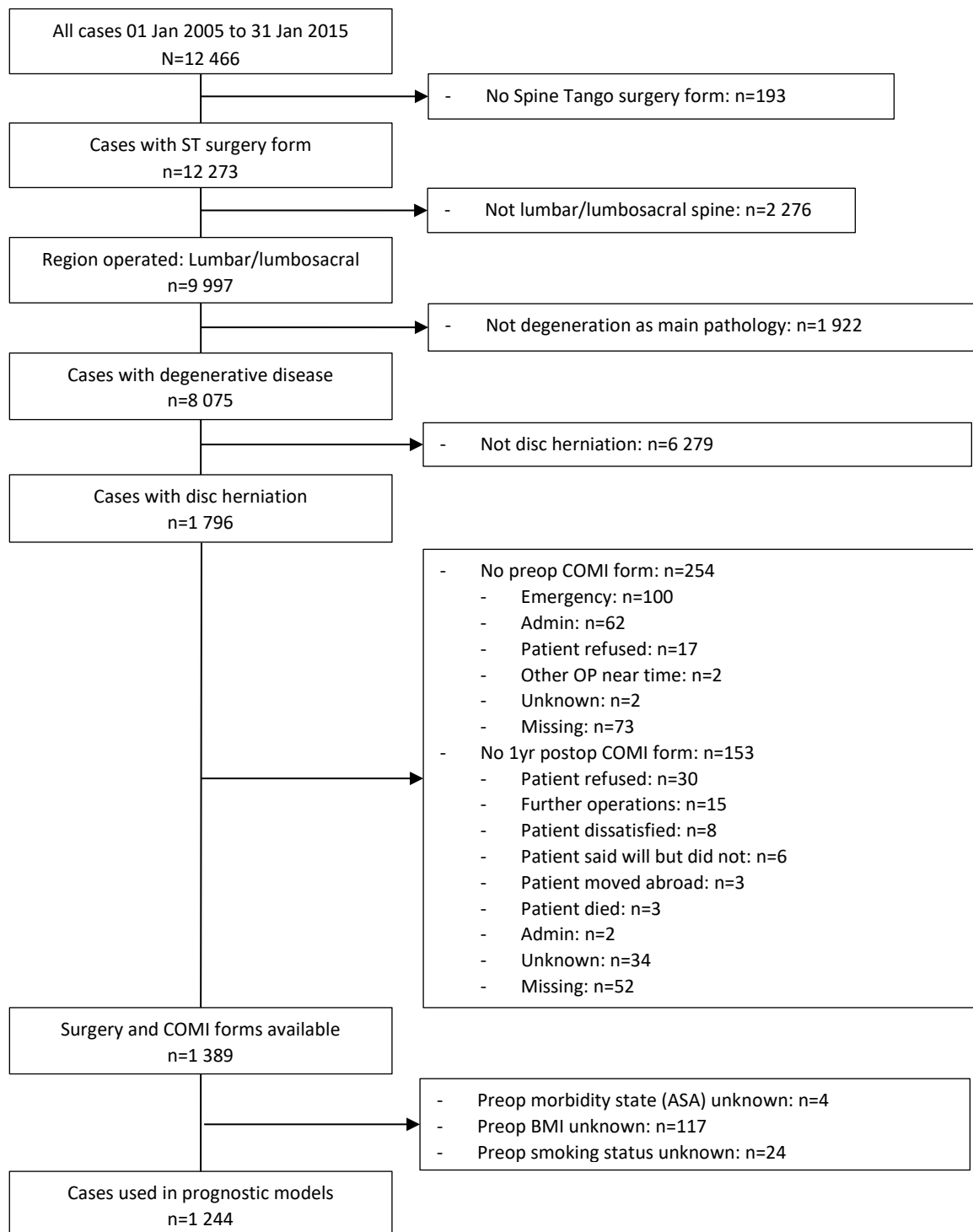


Figure 2

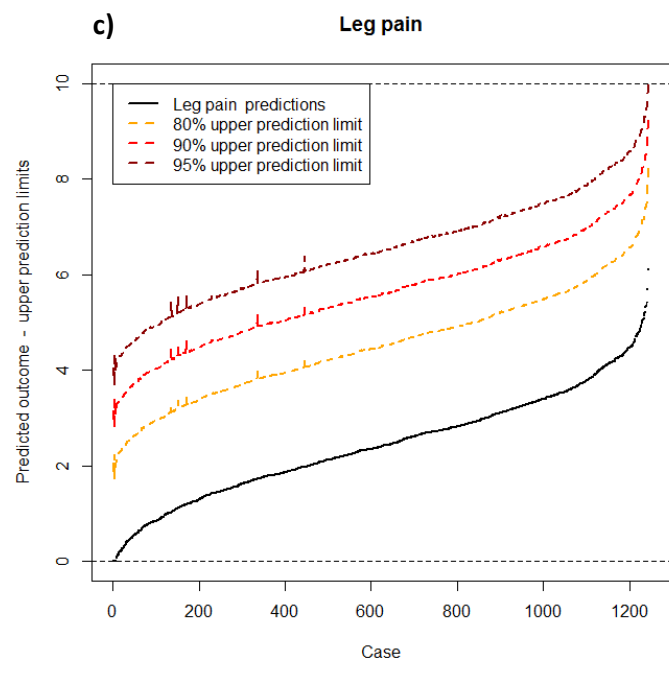
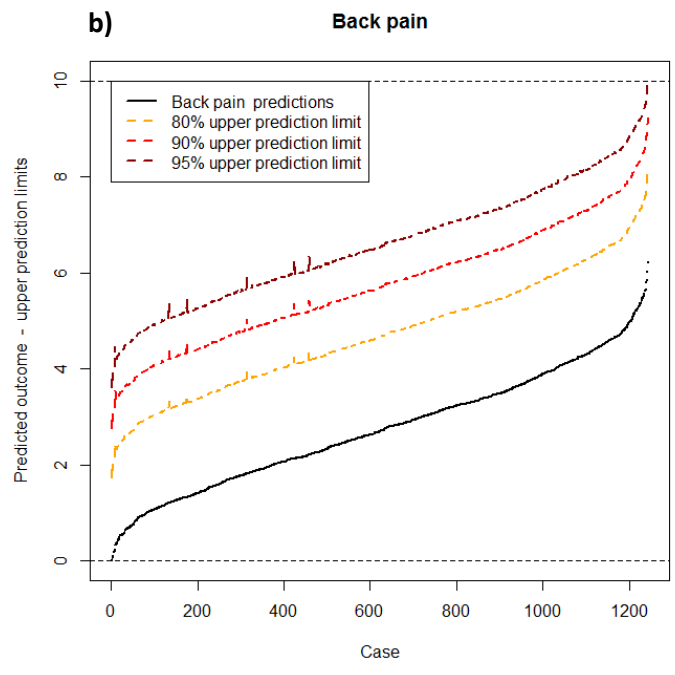
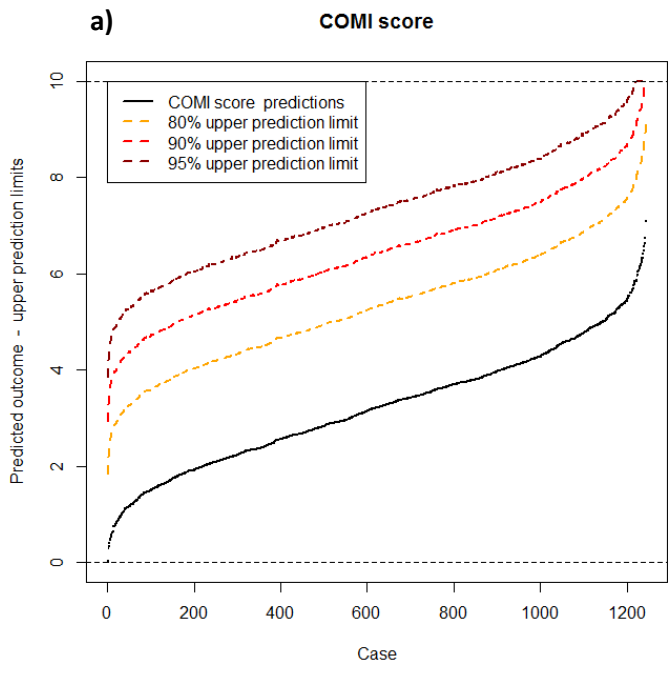
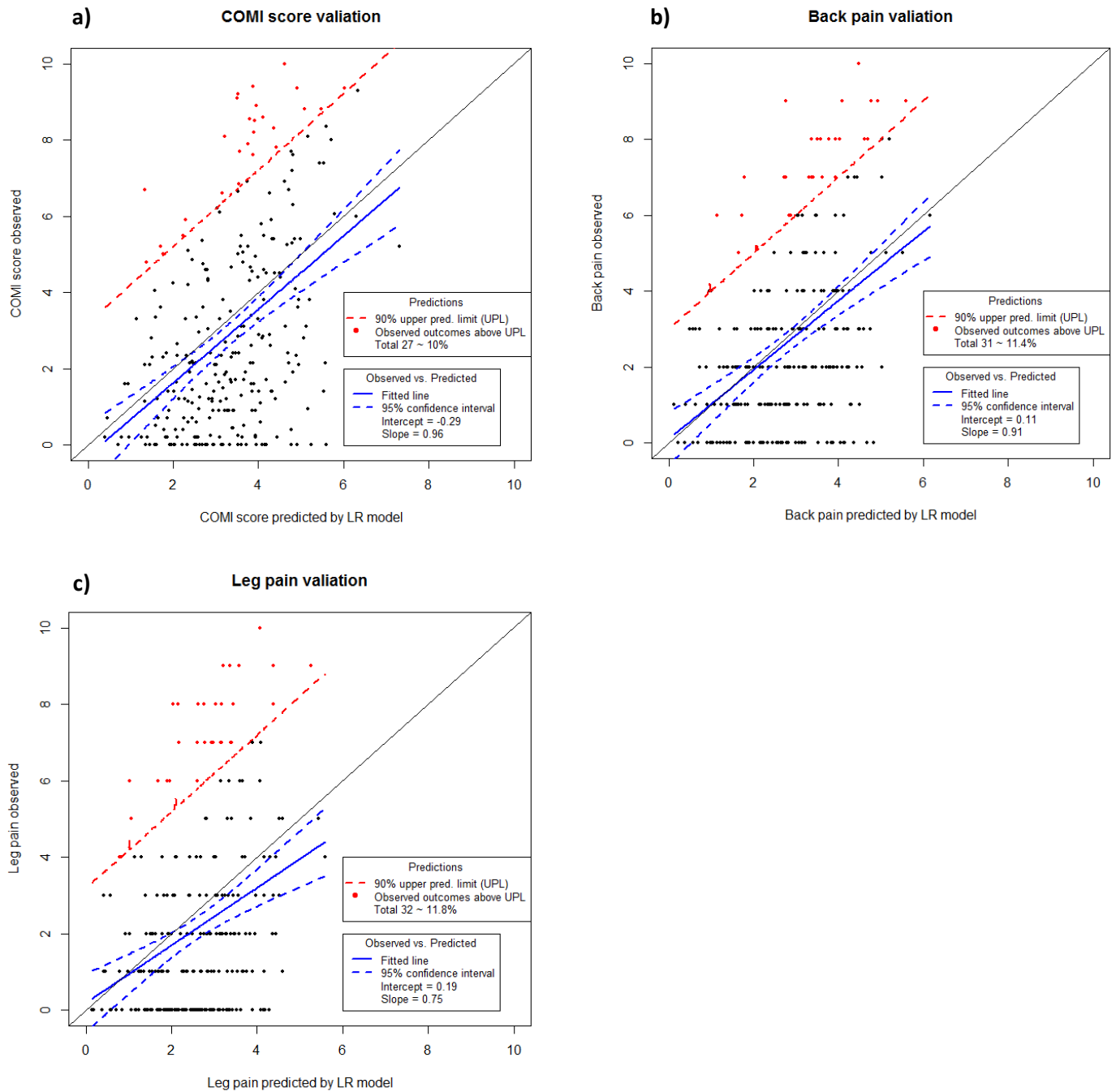


Figure 3



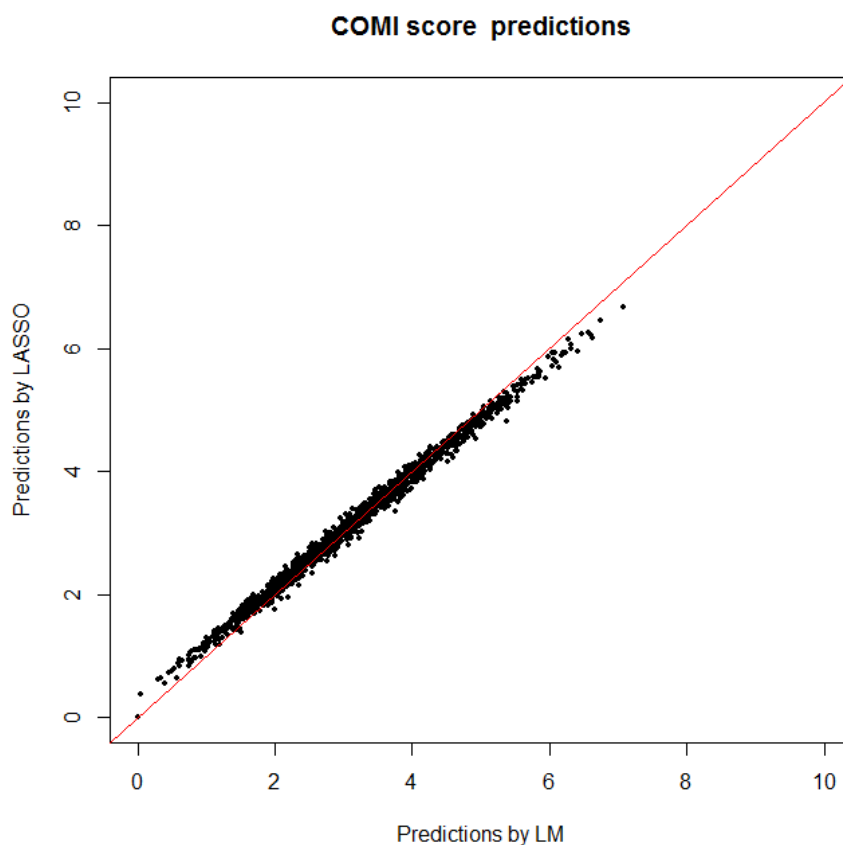
Online Resource 1: Details of LASSO regression modelling

LASSO regression

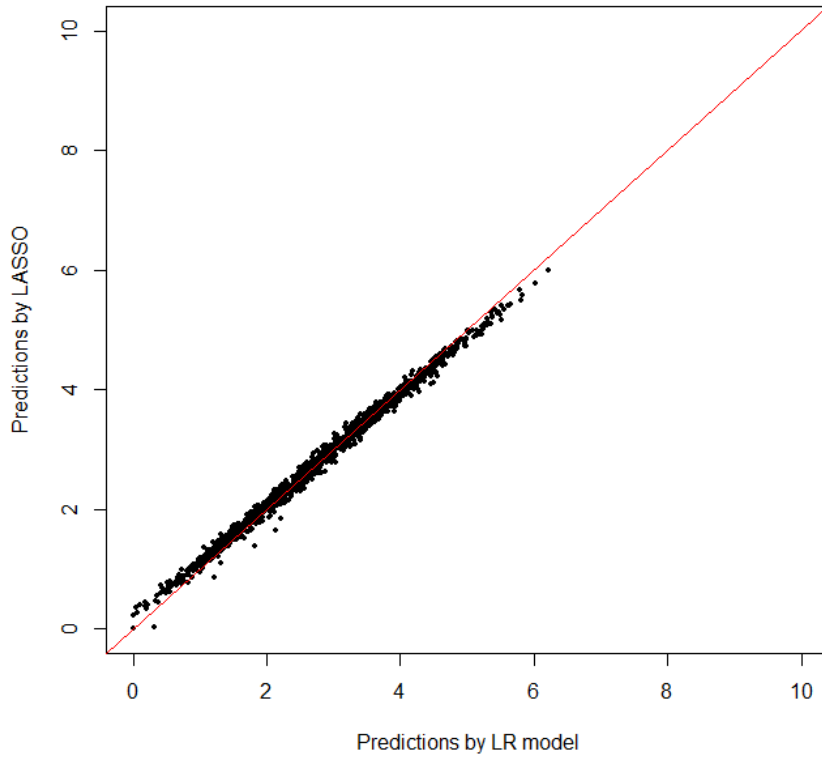
Compared to linear regression, which estimates model coefficients by minimizing the mean squared error, the LASSO uses L_1 penalty for both fitting (minimizing mean absolute error) and penalization of coefficients. As the sum of absolute values of model coefficients is restricted by a constant (lambda), this results in a model with some coefficients set to 0. How many coefficients are set to 0 depends on the parameter lambda. In order to pick the model with the optimal number of predictors, we selected the lambda that leads to the smallest cross-validated mean square error (MSE). When compared to outcomes predicted by the linear regression model, we usually expect a slight attenuation of LASSO predictions towards the mean. We therefore compared the two predictors by plotting them against each other.

Comparison LASSO vs linear regression model

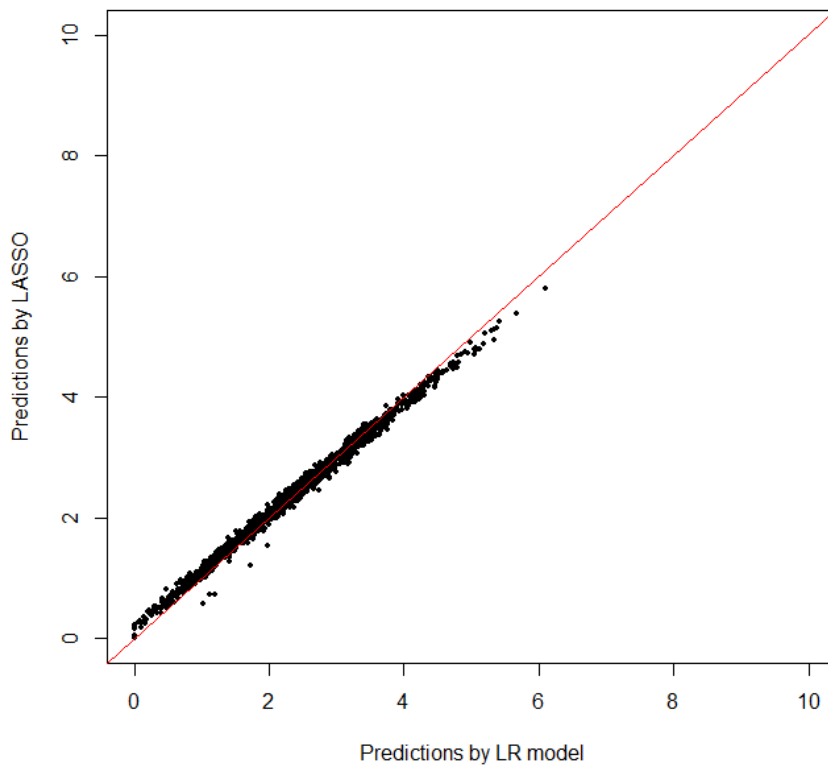
As the figures below show, the predictions obtained with the Lasso regression and the corresponding linear regression models were nearly identical, supporting our approach of using the linear regression models for the subsequent analyses.



Back pain predictions



Leg pain predictions



Online Resource 2: Screenshot and description of online prognostic tool

Online prognostic tool

The figure below shows a screenshot of the current version of the tool. In the left column, patients and surgeons enter the preoperative input variables based on their individual characteristics. The output on the right side then displays the preoperative COMI, back and leg pain scores (red bar), and the corresponding estimated values one year after surgery (green bar). The light green shaded area represents the area below the 90% upper prediction limit, which can be easily communicated to patients. In the example in the figure, the preoperative leg pain of 7 points is expected to improve to 1.9 points; however, with a 90% upper prediction limit of 5.4, the interpretation conveyed to patients is that one out of ten patients with similar characteristics may still have a score higher than this upper bound of 5.4.

