Prediction of outcome after spinal surgery—using The Dialogue Support based on the Swedish national quality register

Peter Fritzell^{1,2} · Johan Mesterton³ · Olle Hagg⁴

Received: 20 May 2021 / Revised: 21 October 2021 / Accepted: 15 November 2021 / Published online: 27 November 2021
© The Author(s) 2021

Abstract

Purpose To evaluate the predictive precision of the Dialogue Support, a tool for additional help in shared decision-making before surgery of the degenerative spine.

Methods Data in Swespine (Swedish national quality registry) of patients operated between 2007 and 2019 found the development of prediction algorithms based on logistic regression analyses, where socio-demographic and baseline variables were included. The algorithms were tested in four diagnostic groups: lumbar disc herniation, lumbar spinal stenosis, degenerative disc disease and cervical radiculopathy. By random selection, 80% of the study population was used for the prediction of outcome and then tested against the actual outcome of the remaining 20%. Outcome measures were global assessment of pain (GA), and satisfaction with outcome.

Results Calibration plots demonstrated a high degree of concordance on a group level. On an individual level, ROC curves showed moderate predictive capacity with AUC (area under the curve) values 0.67–0.68 for global assessment and 0.6–0.67 for satisfaction.

Conclusion The Dialogue Support can serve as an aid to both patient and surgeon when discussing and deciding on surgical treatment of degenerative conditions in the lumbar and cervical spine.

Level of evidence I.

Keywords National quality register · Spine surgery · Outcome · Prediction · Dialogue support

Introduction

The most important question for the patient to be answered before surgery is: “how much better will I be?” This is often a difficult question for the surgeon to answer reliably, as patient-reported outcome after surgery for degenerative spinal conditions demonstrates major heterogeneity. At follow-up, using Swespine, between 10 and 40% of the patients,

depending on the preoperative diagnosis, still may suffer from some spine-related disability and pain [1].

With the development of national quality surgical spine registers such as the Swedish “Swespine”, surgeons and patients are having access to aggregated outcome data, serving as a rough suggestion of possible achievements for an individual patient after surgery. This is, for example, demonstrated in three international Nordic cooperation studies where data from Swespine, NorSpine (Norway) and Danespine (Denmark) are presented [2–4].

It is obvious, however, that translating these data, which are presented on a group level, to individualized assessment of surgical success, may be difficult. This is because several socio-demographic characteristics and other baseline variables, not always known, will modify the outcome.

In 2013, using Swespine data, an analytic project was initiated, together with Region Stockholm (<https://www.sll.se/om-regionstockholm/Information-in-English1/>). The initial aim was to present case-mix-adjusted outcome data publicly for a fair comparison and benchmarking of individual spine

✉ Peter Fritzell
peterfritzell@gmail.com

¹ Present Address: Futurum Academy for Health and Care, Hus B4, Lanssjukhuset Ryhov, 553 05 Jönköping, Sweden

² RKC Spine Center, Södra Fiskartorpsvägen 15H, 114 33 Stockholm, Sweden

³ Department of Learning, Informatics, Management and Ethics, Medical Management Centre, Karolinska Institute, Stockholm, Sweden

⁴ Spine Center Göteborg, Gruvgatan 6, 421 30 Västra Frölunda, Sweden

centres; <https://varдениsiffror.se/jamfor/kallsystem> (only in Swedish). This led to the development of a tool for prediction of individual outcome after surgery for lumbar and cervical degenerative conditions. In 2017, we could present the Dialogue Support for members of the Swedish Society of Spinal Surgeons, and it was later made publicly available; <http://www.4s.nu/4s-f%C3%B6rening/dialogst%C3%B6d-44852774> (only in Swedish).

The support is an interactive web-based instrument to be used in shared decision-making with the patient when discussing surgery for different spinal disorders. After being translated to English and discussed with the Eurospine board, it was made publicly available at Eurospine Home page in October 2020; <https://app.molnify.com/app/7wqw6owgrznr76bkaqc6l4bs7q>.

In Fig. 1a and b, two examples of prediction of outcome using the Dialogue Support are demonstrated. To the left on the screen picture, the patient's individual values of predictor variables are recorded. The right side demonstrates the predicted outcomes for that patient based on the patient's characteristics combined with algorithms describing the relationship between patient characteristics and outcomes based on large amounts of historical data. The pie chart shows predicted probabilities for the five alternatives of pain change and in the banner at the top of the screen, the predicted pain change and satisfaction with outcome are dichotomously summarized into a percentage of success and satisfaction (for description of outcome variables see Methods). The tool can be used by the reader using the following link to the Eurospine Home page: <http://www.eurospine.org>

Patient-centred outcome prediction is a growing focus in spine research, producing several reports annually. The majority discuss prediction in terms of Patient-Reported Outcome Measures (PROMs) [5–15], a few deals with adverse events [16, 17], length of stay [18], revision surgery [19] or return to work [20]. Among the PROM analyses, the outcome measure is usually dichotomized. The predictive modelling differs, but the most frequently used is based on multivariate logistic regression algorithms [5, 6, 9, 11, 13, 14]. In recent years, machine learning has gained interest as an alternative [9, 12, 21, 22]. Further details of available reports related to degenerative spine surgery are presented in Table 1.

The aim of the current study is to evaluate the predictive precision of the Dialogue Support.

Methods

The dialogue support, www.eurospine.org

The Dialogue Support is predicting outcome 1 year after surgery for patients with selected spinal disorders. The

underlying prediction models used have been trained on a sizable body of data throughout Sweden during a 10-year period and are updated every year. The data quantity thus always includes outcomes no more than 1 year old.

The prediction is demonstrated as a proportion of a specific patient group achieving a certain outcome after surgery and answering PROM after 1 year, here global assessment ("Totally pain free-Much better-Somewhat better-Unchanged-Worse") [23] and satisfaction. Each prediction algorithm (one per diagnostic group) is based on approximately 2000–12,000 individuals, depending on the diagnosis and baseline profile. The included diagnosis groups are lumbar disc herniation (LDH), lumbar spinal stenosis (LSS), lumbar degenerative disc disease (DDD) and cervical radiculopathy (CR), which is caused by either disc herniation or foraminal stenosis.

Data source

Swespine in its current form was started 1998 and to date includes approximately 155 000 operated patients (i.e. index procedures used for predictive evaluation) with degenerative conditions in the lumbar and cervico-thoracic spine. National coverage is 95%, completeness 85% and 1-year follow-up over 70%.

Participants in the current study

Patients with their index surgery between 2007-01-01 and 2019-04-01, and who had one-year follow-up, were included in the analyses, resulting in a total of 87 494 patients: 23 087 with LDH, 51 390 with LSS and 5 872 with DDD 7 154 with CR. All patients have given consent prior to registration in Swespine, including information that their data will be used in clinical studies and that they can withdraw their individual data from the registry at any time. The evaluation procedure follows the TRIPOD recommendations (Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) [24].

All analyses were performed separately for the four different subgroups. Patients with missing data on the outcome variable or any of the explanatory variables were excluded. No imputations for missing data were performed.

Outcome

The two outcome measures used were:

- A. Global Assessment (GA) is an ordinal type Likert scale with six response alternatives; "How is your back/leg pain today as compared to before the surgery?" where 0 represents no back/leg pain before the surgery, 1 = completely pain free, 2 = much better, 3 = somewhat

(a)

(b)

Fig. 1 **a** Exemplifying case: The Dialogue Support used for a woman, 40 years, with a lumbar disc herniation; leg pain for 3 months, NRS (Numeric rating Scale)=7/10, no back pain, quality of life (EQ-5D)=0.3/1.0, function (Oswestry Disability Index)=50/100. She has had no previous spine surgery and is a non-smoker. **Assessment:** Based on the prediction model, there is an 89% probability for

a patient with these characteristics to report a successful outcome (dark and light green sectors added). **b** The “same patient” is being a smoker, has had previous spine surgery, is 60 years old and has also back pain (NRS) 6/10 and comorbidities. **Assessment:** Based on the prediction model, there is a 56% probability for a patient with these characteristics to report a successful outcome

Table 1 Compilation of recently published prediction models

Authors	Authors	Diagnosis	Material (n)	Outcome	Follow-up	Method	Test data set	AUC	c-index	R^2	Calibration
16	Bernstein et al.	PLF	7 192/4 182	Adverse event		Log reg	Int/Ext		0.66/0.69		
17	Han et al.	Spine surgery	157 895	Readmission					0.62/0.65		
5	Hegarty et al.	Discectomy	53	Adverse event		Log reg	Ext	0.7–0.76			
25	Huber et al.	Arthroplasty	130 945	Post-op pain	12 m	Log reg	Int		0.66		
6	Janssen et al.	Fusion	202/251	EQ-VAS (dic)	12 m	Machine learning	Ext	0.59–0.87			
7	Karhade et al.	Stenosis	906	Pain/dic	12 m	Log reg	Ext	0.68			
18	Karnuta et al.	Fusion	38 070	function/pain (dic)	12 m	5 different	Ext		0.79/0.69		
8	Khan et al.	Cerv. Myelopathy	93	LOS		Bayes machine learning	Ext	0.94			
9	Khor et al.	Spine surgery	306/616/355	SF36 (PCS) (dic)	12 m	7 machine learning	Ext	0.59–0.73			
19	Lubelski et al.	Cervical surgery	492	Reop	12 m	Log reg?	Ext		69/0.69/0.79		
10	McGirt et al.	Lumbar surgery	1 803	ODI	12 m	Log/Lin reg	Int		0.91		0.47
11	McGirt M et al.	Lumbar surgery	7 618	ODI/EQ/BP/LP (dic)	12 m	Lin reg					
20	Asher et al.	Lumbar surgery	4 694	RTW	3 m	Log reg	Ext		0.69/0.69/0.67/0.64		
12	Merali et al.	Cerv myelopathy	583	SF6D/mJOA (dic)	12 m	Cox reg	Int		0.71		
13	Quddus et al.	Lumbar fusion	100	ODI/LP/BP (dic)	12 m	Machine learning	Ext	0.7/0.73			
14	Rundell et al.	Lumbar surgery	5 840	ODI/LP/BP (dic)	12 m	Log reg?	Ext	0.71/0.72/0.83			
21	Siccoli et al.	Stenosis	635	ODI (dic?)	12 m	Log reg?	Ext		0.66–0.77		
22	Staartjes et al.	Discectomy	422	ODI/LP/BP (dic)	12 m	Machine learning	Int	0.68/0.79/0.72			
				ODI (dic)	12 m	Machine learning/Log reg	Ext	0.87/0.78			
15	Staub et al.	Discectomy	1244 + 364	BP(dic)				0.90/0.55			
				LP(dic)				0.84/0.72			
				COMI/BP/LP	12 m	Lin reg	Ext				Plots

PLF Posterolateral fusion; EQ-VAS Euroqol; Vas, diuc Dichotomized; ODI Oswestry Disability index; SF36 Short Form 36; LÖP Leg pain; BP Back pain; LOS Length of stay; COMI Core Outcome Measure Index; Log reg Logistic regression; Lin reg Linear regression; AUC Area under the curve; Int Internal; Ext External

better, 4 = unchanged, 5 = worse [23]. Patients responding with the 0-alternative are excluded in the analyses. For the cervical spine, the question relates to neck/arm pain. Leg pain is the outcome in Spinal stenosis and LDH groups, back pain in the DDD group and arm pain in the CR group. In the summarized presentation, GA was dichotomized into success (response alternatives 1 and 2) and failure (response alternatives 3–5).

B. Satisfaction (SAT) with treatment outcome, an ordinal Likert scale with three response alternatives (satisfied, hesitant and dissatisfied). In the analysis this variable was dichotomized to satisfied and not satisfied (hesitant or dissatisfied).

Predictors

A large number of socio-demographic and clinical baseline factors deemed relevant for predicting outcome were evaluated for inclusion in the models. Eligible predictors were all demographic and baseline data in Swespine. Link (in Swedish); <http://www.4s.nu/swespine-formul%C3%A4r-44871294>.

Statistical analysis

Development of algorithms underlying the prediction model was carried out in three steps.

1. Model development and variable selection

For the analysis of GA, an ordered probit model was estimated to account for the five levels in the outcome variable. The analysis of satisfaction was based on a logistic regression model. Backward variable selection based on Akaike information criterion was employed to select variables for inclusion in the final model. Model selection was performed out of sample by randomly splitting the data into a training data set (80% of sample) and a test data set (20% of sample).

2. Model validation

To evaluate accuracy of the model, calibration plots and receiver operating characteristic (ROC) curves were used

[25]. In the calibration plots, patients were divided into 20 equally sized groups based on predicted value for their outcome, and then, the average actual outcome within each group was calculated. Diagnostic ability was also evaluated using receiver operating characteristic (ROC) curves, where the ability to classify patients into the correct group is assessed when varying the discrimination threshold. The dotted line indicates that the model predicts no better than chance, described by an area under the curve (AUC/c-statistic) of 0.5. A line extending to the upper left corner indicates perfect separation with an AUC/c-statistic of 1.0. In these analyses, GA was dichotomized into success (response alternatives 1, 2) and failure (response alternatives 3, 4, 5)

3. Model estimation on entire sample

Once variable selection had been performed in step 1 and validated in step 2, the final models were re-estimated on the entire data set, i.e. both the training data and the test data, to get the largest possible number of observation for the final model parameter estimation.

A separate model is estimated for each of the 4 diagnoses of patients and for each of the 2 outcomes, resulting in 8 ROC and 8 calibration plots.

Data management and statistical analysis were conducted using R Statistical Software version 4.0.0.

Results

Description of the study population

The eligible study population was reduced because of drop-outs at follow-up and missing data of predictor variables as shown in Table 2. In general, the differences of baseline data between the diagnostic groups were small or moderate. The DDD group had the longest duration of pain and the highest frequency of earlier spine surgery (Table 3).

Assessment of predictive ability

I. Calibration plots

Table 2 Study population stratified by diagnosis group and outcome measure

Patient group	LDH		CSS		CR		DDD	
	GA	SAT	GA	SAT	GA	SAT	GA	SAT
Patients fulfilling inclusion criteria	23,078	23,078	51,390	51,390	7154	7154	5872	5872
Excluded due to missing data on outcome	7773	7660	13,972	11,779	2524	2138	1509	1499
Excluded due to missing data on predictor(s)	5734	5697	15,731	17,089	1263	951	996	930
Patients in final analysis	9571	9721	21,687	22,522	3367	4065	3367	3443

Table 3 Descriptive statistics of the study population

	Global assessment				Satisfaction			
	LDH	LSS	DDD	CR*	LDH	LSS	DDD	CR*
	Proportion/mean (sd)	Proportion/mean (sd)	Proportion/mean (sd)	Proportion/mean (sd)	Proportion/mean (sd)	Proportion/mean (sd)	Proportion/mean (sd)	Proportion/mean (sd)
Female	0.45	0.52	0.54	0.52	0.45	0.52	0.55	0.50
Smoking	0.12	0.08	0.07	0.15	0.12	0.08	0.07	0.15
Age pension	0.09	0.56	0.04	0.07	0.09	0.57	0.05	0.08
Disability pension	0.23	0.15	0.28	0.32	0.23	0.15	0.28	0.32
Unemployed	0.09	0.10	0.10	0.10	0.09	0.10	0.10	0.11
Previous spine surgery	0.11	0.20	0.30	0.12	0.11	0.19	0.29	0.11
Comorbidity	0.11	0.25	0.12	0.23	0.11	0.25	0.12	0.22
Walking ability < 100 m	0.30	0.34	0.10	0.04	0.30	0.34	0.10	0.04
100–500 m	0.20	0.30	0.17	0.09	0.20	0.30	0.17	0.09
0.5–1 km	0.15	0.16	0.21	0.13	0.15	0.16	0.21	0.13
> 1 km	0.35	0.19	0.52	0.74	0.35	0.20	0.53	0.74
Age	44.6 (13)	65.2 (10.5)	44.7 (10.3)	50.4 (9.2)	44.7 (13.1)	65.5 (10.5)	44.8 (10.4)	50.6 (9.2)
EQ-5D_Base	0.27 (0.34)	0.37 (0.32)	0.35 (0.32)	0.4 (0.32)	0.27 (0.34)	0.38 (0.32)	0.35 (0.32)	0.4 (0.32)
ODI/NDI_Base	47.5 (18.2)	41.9 (15.5)	42.4 (13.7)	44.2 (15.7)	47.5 (18.3)	41.5 (15.6)	42.3 (13.7)	43.1 (16.4)
No back pain	0.06	0.05	0.00	0.02	0.06	0.05	0.00	0.05
Duration back/neck pain < 3 Mths	0.11	0.02	0.00	0.03	0.11	0.02	0.00	0.03
3–12 Mths	0.51	0.20	0.09	0.27	0.50	0.20	0.09	0.27
1–2 Yrs	0.14	0.22	0.17	0.22	0.14	0.22	0.17	0.21
> 2 Yrs	0.18	0.52	0.74	0.47	0.18	0.51	0.74	0.45
VAS back/neck pain	47.4 (28.6)	57.3 (26)	64.5 (19.6)	61.1 (23.4)	47.4 (28.6)	57 (26.2)	64.5 (19.7)	58.5 (25.6)
No leg pain	0.00	0.02	0.23	0.03	0.01	0.03	0.23	0.04
Duration leg/arm pain < 3 Mths	0.16	0.02	0.02	0.04	0.16	0.02	0.02	0.04
3–12 Mths	0.58	0.28	0.14	0.34	0.58	0.28	0.14	0.34
1–2 Yrs	0.14	0.28	0.18	0.26	0.13	0.28	0.18	0.25
> 2 Yrs	0.12	0.40	0.44	0.33	0.12	0.39	0.43	0.33
VAS leg/arm pain	69 (22.6)	66.2 (22.7)	38.4 (28.6)	57.5 (25)	69 (23.1)	64.6 (24)	38.44 (28.67)	56.8 (25.6)
N:o op levels		0.55	0.56			0.55	0.56	
1								
2		0.31	0.40			0.31	0.39	
3		0.11	0.04			0.11	0.04	
4		0.02	0.00			0.02	0.00	
5		0.00	0.00			0.00	0.00	
Completely disappeared	0.37	0.28	0.00	0.19				
Much improved	0.40	0.32	0.00	0.46				
Somewhat improved	0.13	0.18	0.00	0.21				
Unchanged	0.05	0.13	0.00	0.08				
Worse	0.04	0.10	0.00	0.06				
Satisfied					0.79	0.66	0.76	0.73

*CR (Cervical Radiculopathy): Duration back pain corresponds to Duration neck pain, Duration leg pain to Duration arm pain, VAS back pain to VAS neck pain, VAS leg pain to VAS arm pain and ODI to NDI (Neck Disability Index)

The plots demonstrate how well predicted probabilities agree with actual outcome for subgroups with different case mix. Observations were ranked according to the predicted value and grouped in 20 categories in the lumbar diagnostic groups and in 5 categories in the cervical group. The proportion with actual successful outcome (y-axis) and predicted value (x-axis) was calculated for each category and plotted against each other. The solid line represents perfect calibration and dotted line represents the actual results. The concordance between prediction and actual outcome measured with GA for success and satisfaction on a group level was high, with small differences between diagnostic groups. Satisfaction in the DDD group was least concordant. The findings are demonstrated in Fig. 2.

II. ROC curves

The ROC curve demonstrates the ability of the model to separate successful cases and failures. As shown in Fig. 3, the ability of the prediction models to discriminate between successes and failures on individual level is fair, with AUC ranging from 0.67 to 0.68. There were slight differences in model fit between the diagnostic groups. For satisfaction, the AUC value varied more between diagnostic groups, from 0.6 for DDD to 0.67 for CR.

III. Model estimation on entire sample

Table 4 presents the effect of each predictor on the two outcomes for the four different patient groups. Indicators of lower socio-economic status, such as smoking, disability pension and unemployment, were consistently associated

with lower satisfaction and less pain improvement. Previous spine surgery was a negative predictor for all diagnostic groups. Short duration of back/neck pain was associated with more pain improvement and higher satisfaction in all diagnostic groups. Short duration of leg/arm pain was associated with more pain improvement and higher satisfaction in the LDH, CSS and CR groups. Age and gender were of minor importance, as was also ODI, whereas a higher quality of life (EQ-5D) at baseline predicted higher satisfaction at follow-up for all but the CR group.

Discussion

The Dialogue Support, based on the national Swedish quality register “Swespine”, presents the predicted outcome after surgery for degenerative spinal disorders. The outcome is presented as a percentage of outcome according to global assessment of pain (GA) and satisfaction with outcome based on the patient’s characteristics combined with algorithms describing the relationship between patient characteristics and outcomes based on large amounts of historical data.

The calibration plots of subgroups demonstrate a high degree of concordance, with minor differences between diagnostic groups. The message to the patient can be expressed as follows: in the group of earlier operated individuals with a similar baseline profile as you, a certain percentage reported after 1 year “complete relief of pain”, another percentage

Fig. 2 Calibration plots of success in **a** the LDH group, **b** the LSS group, **c** the DDD group and **d** the CR group; of satisfaction in **e** the LDH group, **f** the LSS group, **g** the DDD group and **h** the CR group

Fig. 3 ROC curves of satisfaction in **a** the LDH group, **b** the LSS group, **c** the DDD group and **d** the CR group; of success in **e** the LDH group, **f** the LSS group, **g** the DDD group and **h** the CR group

reported “much better”, a third percentage “somewhat better”, etc. This is visualized in the pie diagram.

On an individual level, as estimated with ROC curves, the precision of the predictive model was fair. For global assessment of pain, the AUC value ranged from 0.67 to 0.68. For satisfaction, it ranged from 0.60 (DDD) to 0.66 (LDH). Other reports with PROMs as outcome measure and logistic regression as analytic method describe AUC values and c-indexes ranging from 0.64 to 0.79, mostly tested on single-centre cohorts [5, 6, 9, 11, 14].

In recent years, new computer based analytic methods, often called “machine learning” and “deep learning”, have been proposed as possibly more powerful methods of data acquisition and analysis. The suggested advantage of these techniques appears not to be determined [26]. Reported AUC values or c-indexes with these analytic methods range from 0.59 to 0.90 [8, 12, 21, 22]. Probably the more critical aspect of outcome prediction, and possibilities of improving precision, is related to addition of more predictor variables. This appears to be more important than focusing on analytic techniques, although different analytic models may change the predictive potential of a particular data set, as has been demonstrated on a hip and knee arthroplasty cohort [27].

The limitation of the current model as such is related to the baseline and socio-demographic variables at hand. Evaluation and development of the Dialogue Support is a continuous process. There could be some changes in the surgery techniques, processing, trainings, support, machineries and medicines affecting outcome. This is considered by yearly updating of the database. The yearly updating of the database and introduction of new baseline variables are

also expected to increase the precision of the model on the individual level over time. Dropouts may have a tendency of worse outcome than consenters also in registers [28], so there is a possible risk of overestimating success in the model. However, this is still an unsolved question [29].

A possible selection bias could be caused by the proportion of patients not having their index surgery recorded. However, there are only exceptional occurrences of patients opting out from Swespine. The major cause of loss to registration is deficient routines in some of the participating centres, not dependent on individual patients. In our interpretation, this does not cause any bias in the national register. A second possible cause of selection bias would be loss to follow-up. This has been assessed in a recent publication [28].

When it comes to application of the Dialogue Support in other countries, the Swedish reference database can be a limitation to generalizability. Spine patients in different countries have different cultural and social conditions, which may affect the predictive ability of the model. Interpretation of predictions should be done with this in mind, until validation tests have been performed. Ideally the model should be tested on a national, or other large, database in the country in question.

The strength of the Dialogue Support is related to the large reference population in the national Swespine Registry, which has a coverage of 95%, a completeness of 85% and a one-year follow-up of more than 70%. Thus, the data that the prediction model is based on well represent the “degenerative spine population” in Sweden over the last ten years. Predictions can thus be generalized in the

Table 4 Impact of predictors on the two outcomes in the four diagnostic groups

	Global assessment				Satisfaction			
	LDH	CSS	DDD	CR*	LDH	CSS	DDD	CR*
	logOR (se)	logOR (se)	logOR (se)	logOR (se)	logOR (se)	logOR (se)	logOR (se)	logOR (se)
Female		− 0.029 (0.015)	− 0.137 (0.038)		− 0.094 (0.052)	0.11 (0.03)		0.099 (0.076)
Smoking	0.177 (0.034)	0.22 (0.026)			− 0.521 (0.071)	− 0.468 (0.052)	− 0.203 (0.153)	− 0.249 (0.098)
Age pension	0.068 (0.048)		0.279 (0.101)		− 0.351 (0.086)	− 0.112 (0.05)	− 0.785 (0.197)	
Disability pension	0.129 (0.027)	0.086 (0.022)	0.098 (0.044)	0.075 (0.042)	− 0.249 (0.061)	− 0.131 (0.047)	− 0.303 (0.093)	
Unemployed	0.318 (0.04)	0.142 (0.024)	0.268 (0.063)	0.275 (0.062)	− 0.615 (0.079)	− 0.317 (0.047)		− 0.516 (0.111)
Previous spine surgery	0.287 (0.036)	0.338 (0.018)	0.213 (0.042)	0.251 (0.058)	− 0.494 (0.076)	− 0.57 (0.036)	− 0.507 (0.09)	− 0.39 (0.109)
Comorbidity	0.238 (0.036)	0.16 (0.017)	0.175 (0.058)	0.153 (0.047)	− 0.33 (0.074)	− 0.254 (0.034)	− 0.114 (0.123)	
Walking ability 100–500 m*	0.075 (0.034)	0.027 (0.019)	0.033 (0.076)	− 0.04 (0.113)	− 0.149 (0.075)	0.012 (0.037)	− 0.158 (0.155)	− 0.033 (0.196)
0.5–1 km*	0.069 (0.039)	0.025 (0.024)	0.098 (0.075)	− 0.112 (0.108)	− 0.14 (0.083)	0.056 (0.048)	− 0.067 (0.153)	0.282 (0.189)
> 1 km*	0.055 (0.035)	− 0.022 (0.024)	− 0.063 (0.077)	− 0.194 (0.101)	− 0.035 (0.074)	0.21 (0.049)	0.303 (0.15)	0.493 (0.173)
Age	0.003 (0.001)	0.01 (0.001)	− 0.005 (0.002)	− 0.004 (0.002)		− 0.013 (0.002)	0.008 (0.004)	− 0.006 (0.004)
EQ-5D Base	0.089 (0.047)		− 0.133 (0.075)		0.213 (0.094)	0.233 (0.061)	0.683 (0.142)	
ODI/NDI_ Base	0.003 (0.001)	0.007 (0.001)	0.007 (0.002)	0.013 (0.002)		− 0.007 (0.001)		− 0.017 (0.003)
Duration back/neck pain < 3 mths**	− 0.017 (0.066)	− 0.228 (0.076)	0.845 (0.667)	− 0.775 (0.235)	0.42 (0.155)	0.456 (0.166)	0.267 (1.534)	0.897 (0.404)
3–12 mths**	− 0.037 (0.053)	− 0.101 (0.042)	0.978 (0.542)	− 0.363 (0.154)	0.18 (0.133)	0.236 (0.088)	− 0.022 (1.105)	0.815 (0.201)
1–2 yrs**	0.026 (0.063)	− 0.009 (0.041)	1.133 (0.54)	− 0.221 (0.155)	− 0.196 (0.144)	0.126 (0.085)	− 0.462 (1.099)	0.605 (0.201)
> 2 yrs**	0.154 (0.062)	0.08 (0.039)	1.337 (0.539)	− 0.036 (0.153)	− 0.373 (0.142)	− 0.045 (0.082)	− 0.641 (1.096)	0.435 (0.191)
VAS back/neck pain	0.005 (0)	0.004 (0)		0.005 (0.001)	− 0.011 (0.001)	− 0.007 (0.001)		
Duration leg/arm pain < 3 mths**	− 0.19 (0.174)	− 0.386 (0.078)		− 0.18 (0.189)		0.971 (0.148)		1.317 (0.364)
3–12 mths**	− 0.059 (0.171)	− 0.351 (0.056)		− 0.042 (0.123)		0.684 (0.092)		0.664 (0.219)
1–2 yrs**	0.145 (0.174)	− 0.179 (0.055)		0.081 (0.124)		0.433 (0.091)		0.516 (0.219)
> 2 yrs**	0.212 (0.174)	− 0.067 (0.055)		0.094 (0.122)		0.308 (0.09)		0.454 (0.215)
VAS leg/arm pain			0.002 (0.001)	− 0.002 (0.001)	0.005 (0.001)	0.003 (0.001)		− 0.004 (0.002)
N:o op levels 2***		0.05 (0.016)				− 0.088 (0.033)	− 0.155 (0.086)	

Table 4 (continued)

	Global assessment				Satisfaction			
	LDH	CSS	DDD	CR*	LDH	CSS	DDD	CR*
	logOR (se)	logOR (se)	logOR (se)	logOR (se)	logOR (se)	logOR (se)	logOR (se)	logOR (se)
3***		0.055 (0.024)				− 0.056 (0.048)	− 0.445 (0.199)	
4***		0.166 (0.049)				− 0.228 (0.094)	− 0.586 (0.564)	
5***		0.276 (0.159)				− 0.757 (0.318)	10.213 (196.968)	
Constant					1.875 (0.164)	1.734 (0.18)	1.446 (1.117)	0.91 (0.377)
Completely pain free/ much improved	0.364 (0.189)	0.55 (0.086)	0.539 (0.559)	− 0.615 (0.289)				
Much improved/ somewhat improved	1.503 (0.189)	1.439 (0.086)	1.938 (0.56)	0.773 (0.289)				
Somewhat improved/ unchanged	2.128 (0.19)	1.984 (0.087)	2.634 (0.56)	1.522 (0.289)				
Unchanged/ worse	2.61 (0.19)	2.551 (0.087)	3.102 (0.561)	2.063 (0.29)				

For facility of interpretation, model coefficients are presented on the log scale (logs odds ratios). As opposed to the odds scale, the log odds scale is symmetrical around zero. GA is coded so that 1 is best assessment and 5 is worst assessment, while satisfaction is coded so that 1 is satisfied. Hence, a negative/positive coefficient implies that a higher value on the predictor is associated with lower/higher probability of worse global assessment and higher satisfaction. An empty cell implies that the variable was not included in the final model after the variable selection procedure

*CR (cervical radiculopathy): Duration back pain corresponds to Duration neck pain, Duration leg pain to Duration arm pain, VAS back pain to VAS neck pain and VAS leg pain to VAS arm pain and ODI to NDI (Neck Disability Index)

entire nation and applied to all spine centres. Dropouts and missing data may infer a limitation to the generalizability, which we hope to reduce with increasing web-based registration.

In the aggregated perspective, the Dialogue support, acting as “one piece of the puzzle”, can support the clinician’s clinical experience/judgement. In popular terms, using the Dialogue support, it is possible to describe outcome as a certain probability to have a successful and satisfactory outcome after a proposed surgical intervention.

Thus, the Dialogue support offers the opportunity to both patient and surgeon to contemplate and discuss the probability of benefit and risk based on more substantial evidence than the experience and conjecture of an individual surgeon. It is also an interesting opportunity to start international research cooperation based on the Dialogue support.

Our ambition is to validate these models further once new data becomes available (data are continuously collected in Swespine, approximately 10 000 patients per year), and we can also foresee validating models in other

countries where similar (albeit not identical) data are collected.

Conclusion

The Dialogue Support is a useful prediction tool with an accuracy which is high on a group level and moderate on an individual level. It can serve as an aid to both patient and surgeon when discussing a surgical treatment of degenerative conditions in the lumbar and cervical spine.

Acknowledgements We thank all members of the Steering committee of Swespine for collaboration and support necessary for the development of the Dialogue support and the current evaluation: Allan Abbott PT PhD, Carina Blom secretary, Paul Gerdhem MD PhD, Håkan Löfgren MD PhD, Lena Mellgren secr, Catharina Parai MD PhD, Björn Strömqvist MD PhD, Olof Thoreson MD PhD and finally, not the least, David Bergqvist from Ivbar/Logex for his statistical work.

Authors’ contribution All authors have contributed to data collection, analysis and editing of the manuscript.

Funding No funding was obtained.

Data Availability Individual data are not publicly available but on a group level through yearly reports by the steering committee of Swespine (www.4s.se).

Declarations

Conflict of interest JM holds stock in LOGEX Healthcare Analytics, a company specialized in analysis of health care data. The other authors declare no conflict of interest.

Consent for publication All patients in Swespine are given written information about the registry, including information that their data can be used for research and publication if they accept to participate in the register and that they can withdraw their data at any time.

Ethical Approval Retrospective data are used in the current article. All data are anonymized. There are no individual data in the Dialogue support, but all data are on group level. Data are national and cannot be traced to specific location or to an individual, and are not included in the medical record. According to Swedish law, this implies that ethical approval is not needed.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Fritzell P, Hagg O, Gerdhem P, Abbott A, Parai C, Thoreson O, Stromqvist B, Mellgren L, Blom C (2018) Swespine 25 years. 2018 annual report follow up of spine surgery performed in sweden in 2017. ISBN: 978-91-983912-3-7
- Lagerbäck T, Fritzell P, Hägg O, Nordvall D, Lønne G, Solberg TK, Andersen MØ, Eiskjær S, Gehrchen M, Jacobs WC, van Hooff ML, Gerdhem P (2019) Effectiveness of surgery for sciatica with disc herniation is not substantially affected by differences in surgical incidences among three countries: results from the Danish, Swedish and Norwegian spine registries. *Eur Spine J* 11:2562–2571
- Lønne G, Fritzell P, Hägg O, Nordvall D, Gerdhem P, Lagerbäck T, Andersen M, Eiskjær S, Gehrchen M, Jacobs W, van Hooff ML, Solberg TK (2019) Lumbar spinal stenosis: comparison of surgical practice variation and clinical outcome in three national spine registries. *Spine J* 1:41–49
- Andersen MØ, Fritzell P, Eiskjær S, Lagerbäck T, Hägg O, Nordvall D, Lønne G, Solberg T, Jacobs W, van Hooff M, Gerdhem P, Gehrchen M (2019) Surgical treatment of degenerative disk disease in three scandinavian countries: an international register study based on three merged national spine registers. *Glob Spine J* 9(8):850–858
- Hegarty D, Shorten G (2012) Multivariate prognostic modeling of persistent pain following lumbar discectomy. *Pain Physician* 15(5):421–434
- Janssen ERC, Punt IM, van Kuijk SMJ, Hoebink EA, van Meeteren NLU, Willems PC (2020) Development and validation of a prediction tool for pain reduction in adult patients undergoing elective lumbar spinal fusion: a multicentre cohort study. *Eur Spine J* 29(8):1909–1916
- Karhade AV, Fogel HA, Cha TD, Hershman SH, Doorly TP, Kang JD, Bono CM, Harris MB, Schwab JH, Tobert DG (2021) Development of prediction models for clinically meaningful improvement in PROMIS scores after lumbar decompression. *Spine J* 21(3):397–404
- Khan O, Badhiwala JH, Witiw CD, Wilson JR, Fehlings MG (2020) Machine learning algorithms for prediction of health-related quality-of-life after surgery for mild degenerative cervical myelopathy. *Spine J* 8: S1529–9430(20)30047–4
- 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
- McGirt MJ, Sivaganesan A, Asher AL, Devin CJ (2015) Prediction model for outcome after low-back surgery: individualized likelihood of complication, hospital readmission, return to work, and 12-month improvement in functional disability. *Neurosurg Focus* 39(6):E13
- McGirt MJ, Bydon M, Archer KR, Devin CJ, Chotai S, Parker SL, Nian H, Harrell FE Jr, Speroff T, Dittus RS, Philips SE, Shafrey CI, Foley KT, Asher AL (2017) An analysis from the quality outcomes database, Part 1. Disability, quality of life, and pain outcomes following lumbar spine surgery: predicting likely individual patient outcomes for shared decision-making. *J Neurosurg Spine* 27(4):357–369
- Merali ZG, Witiw CD, Badhiwala JH, Wilson JR, Fehlings MG (2019) Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy. *PLoS ONE* 14(4):e0215133
- Quddusi A, Eversdijk HAJ, Klukowska AM, de Wispelaere MP, Kernbach JM, Schröder ML, Staartjes VE (2020) External validation of a prediction model for pain and functional outcome after elective lumbar spinal fusion. *Eur Spine J* 29(2):374–383
- Rundell SD, Pennings JS, Nian H, Harrell FE Jr, Khan I, Bydon M, Asher AL, Devin CL, Archer KR (2020) Adding 3-month patient data improves prognostic models of 12-month disability, pain, and satisfaction after specific lumbar spine surgical procedures: development and validation of a prediction model. *Spine J* 20(4):600–613
- Staub LP, Aghayev E, Skrivankova V, Lord SJ, Haschtmann D, Mannion AF (2020) Development and temporal validation of a prognostic model for 1-year clinical outcome after decompression surgery for lumbar disc herniation. *Eur Spine J* 29(7):1742–1751
- Bernstein DN, Keswani A, Chi D, Dowdell JE, Overley SC, Chaudhary SB, Mesfin A (2019) Development and validation of risk-adjustment models for elective, single-level posterior lumbar spinal fusions. *J Spine Surg* 5(1):46–57
- Han SS, Azad TD, Suarez PA, Ratliff JK (2019) A machine learning approach for predictive models of adverse events following spine surgery. *Spine J* 19(11):1772–1781
- Karnuta JM, Golubovsky JL, Haeberle HS, Rajan PV, Navarro SM, Kamath AF, Schaffer JL, Krebs VE, Pelle DW, Ramkumar PN (2020) Can a machine learning model accurately predict patient resource utilization following lumbar spinal fusion? *Spine J* 20(3):329–336
- 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

20. Asher AL, Devin CJ, Archer KR, Chotai S, Parker SL, Bydon M, Nian H, Harrell FE Jr, Speroff T, Dittus RS, Philips SE, Shaffrey CI, Foley KT, McGirt MJ (2017) An analysis from the quality outcomes database, Part 2. Predictive model for return to work after elective surgery for lumbar degenerative disease. *J Neurosurg Spine* 27(4):370–381
21. Siccoli A, de Wispelaere MP, Schröder ML, Staartjes VE (2019) Machine learning-based preoperative predictive analytics for lumbar spinal stenosis. *Neurosurg Focus* 46(5):E5
22. Staartjes VE, de Wispelaere MP, Vandertop WP, Schröder ML (2019) Deep learning-based preoperative predictive analytics for patient-reported outcomes following lumbar discectomy: feasibility of center-specific modeling. *Spine J* 19(5):853–861
23. Parai C, Hägg O, Lind B, Brisby H (2018) The value of patient global assessment in lumbar spine surgery: an evaluation based on more than 90,000 patients. *Eur Spine J* 27(3):554–563
24. Collins GS, Reitsma JB, Altman DG, Moons KGM (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement *BMC Medicine* 13:1
25. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Pencina MJ, Kattan MW (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 21(1):128–138
26. <https://thestatsgeek.com/2019/08/08/whats-the-difference-between-statistics-and-machine-learning/>
27. Huber M, Kurz C, Leidl R (2019) Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning. *BMC Med Inform Decis Mak* 19(1):3
28. Parai C, Hägg O, Willers C, Lind B, Brisby H (2020) Characteristics and predicted outcome of patients lost to follow-up after degenerative lumbar spine surgery. *Eur Spine J* 29(12):3063–3073
29. Solberg TK, Sørli A, Øystein KS, Nygaard P, Ingebrigtsen T (2011) Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? A study of responding and non-responding cohort participants from a clinical spine surgery registry. *Acta Orthop* 82(1):56–63

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.