BMC Musculoskeletal Disorders

STUDY PROTOCOL

Open Access

The strategic use of Big Data - A study protocol for a multicenter clinical trial testing if the use of the Swespine Dialogue Support alter outcomes in degenerative spine surgery

Eric Brisby Enger[1*], Ludvig Valentin-Askman[2], Olle Hägg[3], Peter Fritzell[4,5] and Catharina Parai[6,7]

Abstract

Background Patients surgically treated for lumbar spinal stenosis or cervical radiculopathy report improvement in approximately two out of three cases. Advancements in Machine Learning and the utility of large datasets have enabled the development of prognostic prediction models within spine surgery. This trial investigates if the use of the postoperative outcome prediction model, the Dialogue Support, can alter patient-reported outcome and satisfaction compared to current practice.

Methods This is a prospective, multicenter clinical trial. Patients referred to a spine clinic with cervical radiculopathy or lumbar spinal stenosis will be screened for eligibility. Participants will be assessed at baseline upon recruitment and at 12 months follow-up. The Dialogue Support will be used on all participants, and they will thereafter be placed into either a surgical or a non-surgical treatment arm, depending on the decision made between patient and surgeon. The surgical treatment group will be studied separately based on diagnosis of either cervical radiculopathy or lumbar spinal stenosis. Both the surgical and the non-surgical group will be compared to a retrospective matched control group retrieved from the Swespine register, on which the Dialogue Support has not been used. The primary outcome measure is global assessment regarding leg/arm pain in the surgical treatment group. Secondary outcome measures include patient satisfaction, Oswestry Disability Index (ODI), EQ-5D, and Numeric Rating Scales (NRS) for pain. In the non-surgical treatment group primary outcome measures are EQ-5D and mortality, as part of a selection bias analysis.

Discussion The findings of this study may provide evidence on whether the use of an advanced digital decision tool can alter patient-reported outcomes after surgery.

Trial registration The trial was retrospectively registered at ClinicalTrials.gov on April 17th, 2023, NCT05817747.

Protocol version 1.

Trial design Clinical multicenter trial.

Keywords Spine, Prediction modelling, Prediction of Outcome, Decision making

*Correspondence:
Eric Brisby Enger
ericbrisbyenger@gmail.com

Full list of author information is available at the end of the article

Introduction

Advancements in computational power has enabled the application of numerous Machine Learning (ML) based prediction models within the field of degenerative spine surgery [1, 2]. Most of these have undergone internal validation, however few are externally validated. ML is a domain within artificial intelligence (AI) that uses algorithms and input data to make a prediction or a classification [3].

Data in health care registers provide a ground for further development of personalized predictive models in medicine [4]. These big data sets can be used in machine learning and statistical modeling to identify patterns and relationships that might aid in predicting outcome [5].

The Swespine register features about 155 000 index surgeries. Degenerative lumbar surgery account for 85%. The national coverage is 97%, completeness about 85%, and one-year follow-up rate is just below 70% [6]. The register uses a standardised form to collect the data from all participating centres.

The Dialogue Support is a web-based tool based on Swespine data (Fig. 1). Its purpose is to guide during the meeting between spine surgeons and patients leading to the shared decision making whether to opt for surgical treatment or not. The Dialogue Support is a patient-specific prediction tool constructed with the help of a logistic regression model that predicts postoperative outcome at one year after surgery based on patient characteristics, which has been shown to alter outcome after spinal surgery (Table 1) [7]. The Dialogue Support offers a prediction algorithm for cervical radiculopathy (CR) and lumbar spinal stenosis (LSS) [8]. The predictions are presented in two patient-reported outcome measures

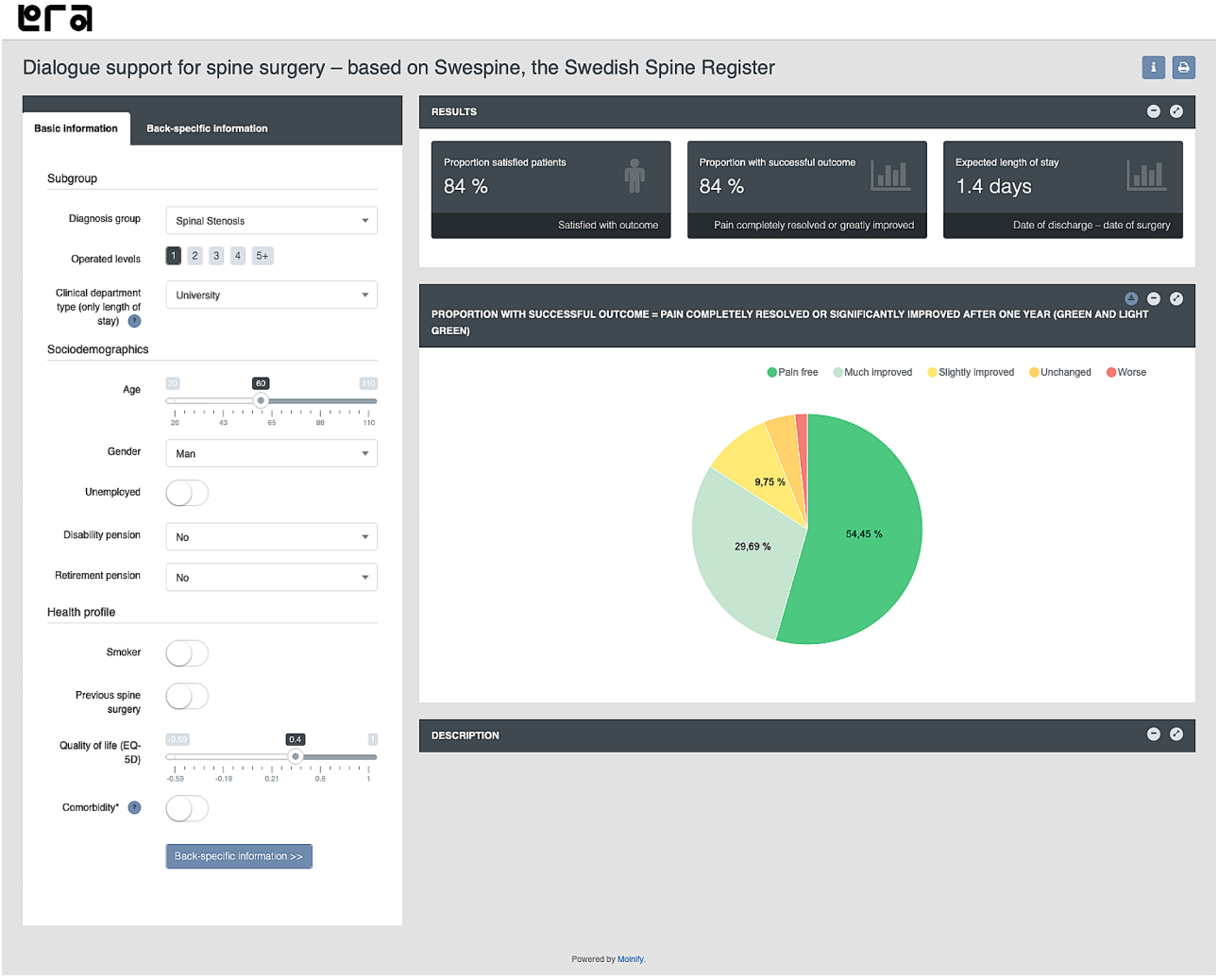

Fig. 1 The prediction model Dialogue Support as it appears on-line. The patient characteristics are filled out under "Basic information" and "Back-specific" information on the left, and the prediction of outcome is visualized on the right. Screenshot from the Dialogue support (available at: https://app.molnify.com/app/7wqw6owgrznr76bkaqc6l4bs7q)

Table 1 Independent variables in the swespine dialogue support, which also serve as baseline characteristics

Clinical setting	University hospital/Private hospital
Age	Years
Sex	Man/Woman
Not employed	Yes/No
Sick pension	Yes/No
Retirement	Yes/No
Smoker	Yes/No
Previous spine surgery	Yes/No
Comorbidity	Yes/No
EQ-5D index	-0.59-0.99
Walking distance	More than 1 km/0.5–1 km/100–500 m/0–100 m
Pain duration leg	No pain/Less than 3 months/3–12 months/1–2 years/ more than 2 years
Pain duration back	No pain/Less than 3 months/3–12 months/1–2 years/ more than 2 years
Preoperative pain leg	NRS 0–10
Preoperative pain back	NRS 0–10
ODI/NDI	0-100/0–50

Oswestry Disability Index (ODI). Euroqol-5-Dimensions quality of life questionnaire (EQ-5D). Numeric Rating Scales (NRS). Neck disability index (NDI). Comorbidity: Heart disease, neurological disease, cancer, other disease affecting walking distance and other disease giving pain.

(PROM): Global Assessment (GA) [9] and Patient Satisfaction.

LSS is a degenerative condition where changes in the discs, ligamentum flavum and facets joints causes a narrowing of the lumbar spinal canal which may result in, pseudo-claudication, and or radiculopathy in the lower extremities [10]. CR is also caused by degenerative changes leading to compression of one or several nerve roots in the cervical spine with clinical manifestations including pain, sensory or motor deficits and diminished reflexes in the upper extremities [11].

The predictive precision of the Dialogue Support was moderate on the individual level, as demonstrated with an Area Under Curve (AUC) ranging from 0.67 to 0.68, in the validation data set during development [8]. An external validation on Danespine register data found similar AUC values for LSS [12]. The usability in clinical practice for the proposed models are yet to be tested.

Less than 2/3 of patients undergoing spinal surgery report significant improvement with postoperative outcome according to GA [6]. This relatively low success rate may be attributed to factors such as comorbidity, age, and socioeconomics, but also reflects the difficulty in assessing which patients will benefit from surgery and who will not. The aim of this study is to determine if the use of the Dialogue Support alters patient-reported outcome compared to current practice in patients diagnosed with lumbar spinal stenosis or cervical radiculopathy.

Methods

Study setting

A prospective, multicenter clinical trial where participants will be assessed at baseline upon recruitment and at one-year follow-up. The Dialogue Support will be used on all participants when surgeon and patient decide upon either surgical or non-surgical treatment. The CR and LSS patients will be collected and compared separately, see Fig. 2. The one-year postoperative outcome of the surgical group will be compared to the outcome of a retrospective control group consisting of previously operated individuals from before the introduction of the Dialogue Support. The non-surgical group will be compared to a cohort of previously operated individuals from before the introduction of the Dialogue Support regarding quality of life, to investigate if the Dialogue Support deselects patients that would have benefitted from an operation.

Eligibility criteria

The patients are recruited from eight different clinics, two university hospitals and six private clinics. In Sweden, most of the routine surgery on patients with low comorbidity is conducted in private clinics, while patients with more advanced disease spectra are operated in hospitals with access to intensive care units, i.e., university hospitals [6].

All patients above 18 years of age, who are referred to one of the participating clinics regarding surgery for LSS or CR with an MRI examination within the last 12 months confirming the diagnosis, concurrent with symptoms described by the referring physician, will be screened for eligibility. The patients will receive a written study information by mail, an informed consent form, and a request to kindly arrive at the outpatient clinic approximately 30 min before the doctor appointment to have sufficient time to fill out the baseline questionnaire needed to run the Dialogue Support algorithm (Table 1).

At the end of the appointment the surgeon will decide if the patient is to be included in the study or not relative to the inclusion and exclusion criteria. Excluded patients

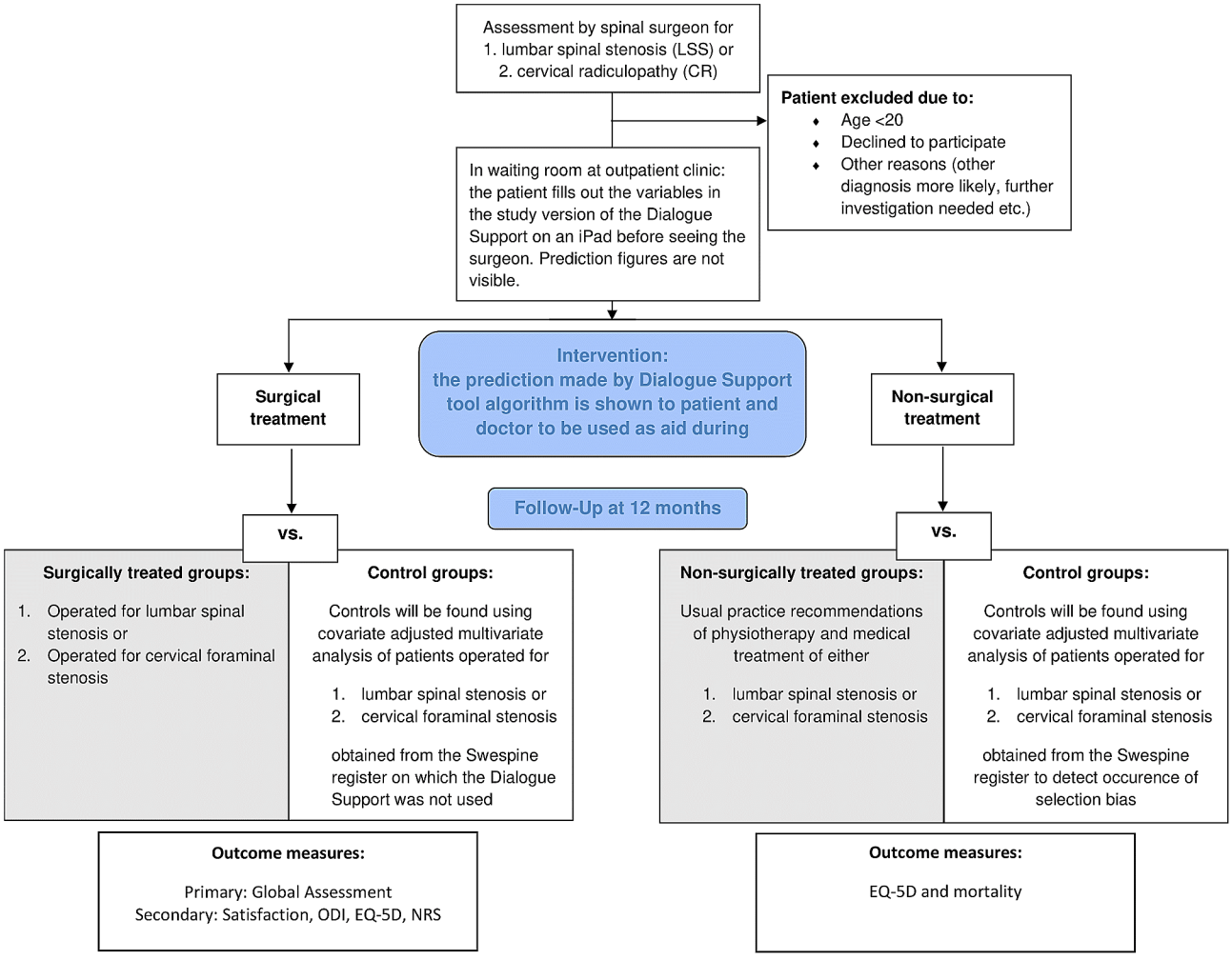

Fig. 2 Study flow chart

will be informed on their exclusion from the study as well as the rationale for exclusion. All data collected on excluded patients will be deleted from the database and will not be used for any analysis.

Participants will be excluded if another condition is found to be the reason for the symptoms, or if there is a need for further investigation that prevents inclusion within one month of the visit at the outpatient clinic. This time limit was set due to practical reasons.

Interventions

If the patient is included the Dialogue Support is used as a complement to routine practice as the surgeon informs about surgical and non-surgical treatment options. The patient will thereafter be put into either of the two treatment arms (i.e., surgical, or non-surgical treatment) according to what has been agreed upon between spine surgeon and patient (Fig. 1). The non-surgical group will follow the usual care recommendations including physiotherapy and other pain-relieving treatment.

The non-surgical group will receive a follow-up questionnaire one-year after inclusion and the surgical group will be followed-up one-year post-surgery. One reminder will be sent.

The control group will consist of patients registered in Swespine within the last five years (2014–2018), before the Dialogue Support was made publicly available. The patient-reported outcomes of surgery at the one-year follow-up for LSS and CR respectively have been stable during the last decade and the use of a retrospective control group is therefore considered acceptable [6].

Outcome variables

Surgical treatment group

The primary outcome measure is GA, which was used for the power calculation. GA is a retrospective single-item question ("How is your back/leg or neck/arm pain today as compared to before the surgery?"), with six response options (no leg pain before surgery/completely gone/much better/slightly better/unchanged/worse) [9]. GA

has been found to be equivalent to multi-item PROMs [13]. In the analyses, GA will be dichotomized into successful outcome (completely gone/much better) and not successful outcome (slightly better/unchanged/worse). The percentage of successful outcome in the surgery group and the control group will be compared.

Secondary outcomes are Satisfaction, the Oswestry Disability Index (ODI), the Neck Disability Index (NDI), the Euroqol-5-Dimensions quality of life questionnaire (EQ-5D), Numeric Rating Scales for back/neck and leg/arm pain respectively (NRS), and mortality.

Like GA, the variable Satisfaction is also a retrospective question ("How is your attitude to the results of your undertaken back/neck surgery?") using a Likert-scale with the response options (satisfied uncertain/dissatisfied). ODI assesses the impact of lumbar pain on a patient´s physical function. It consists of ten questions concerning intensity of pain, ability to care for oneself, lifting, ability to walk, ability to sit, ability to walk, sexual function, social life, sleep quality and ability to travel [14]. ODI is one of the most common outcome variables regarding back pain. NDI is used to assess the self-rated disability in patients with neck pain consisting of ten questions scaled from 0 to 5 concerning pain-intensity, ability to care for oneself, lifting, reading, headache, concentration, working, driving, sleep and recreational activity. NDI is one of the most common outcome variables in studies on cervical radiculopathy [15]. The EQ-5D is a validated instrument to assess quality of life [16]. It describes health in 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The EQ VAS scale measures health state on a scale from 0 to 100. The EQ-5D is presented either as an index or by each separate dimension. The NRS is a common way of grading pain intensity and is presented to the patient as a line with numbers from 0 to 10 and the patient is asked to grade their pain intensity where zero is no pain and ten is maximal pain [17].

Minimal Important Change (MIC) is the minimal level of change that is considered meaningful to patients. MIC-values were previously calculated for NRS (NRS_{back} 2.7, NRS_{leg} 3.2, NRS_{neck} 1.2 and NRS_{arm} 1.3), ODI [18] and NDI [3, 5, 18–20]. The use of MIC estimate for EQ-5D is discouraged as the measurement error (i.e., the smallest detectable change) has been found to be considerably larger than the MIC value [19].

Non-surgical treatment group

The outcome of the non-surgical group will be studied for selection bias purposes. There is a risk that patients with a predicted low probability of a successful outcome of surgery will be allocated to the non-surgical group even though he or she might have been better off with surgery. Therefore, EQ-5D and mortality will be assessed and compared to a previously surgically treated cohort with similar baseline characteristics. The choice of EQ-5D assumed that quality of life would be the most important aspect for this group of participants. Since MIC estimates for the EQ-5D are unreliable, group comparisons will be made by reporting on changes in index-values and in EQ-VAS scores between baseline and follow-up. Changes in dimension levels will be visualized in bar charts.

Mortality at one-year follow-up will also be estimated in these groups because a higher level of mortality in the non-surgical group compared to its surgically treated control group could indicate that surgeons may have chosen not to operate patients, they considered to be frailer.

Participant timeline

A visit at the spine surgeon outpatient clinic is approximately 30 min. Since approximately 15 min is required for the patient to fill out the baseline variables needed to run Dialogue Support algorithm, that is performed independently ahead of the visit with the surgeon to avoid risk of performance bias. Hence, during 2021, a study database was constructed enabling the baseline variables - filled out by each participant right before the appointment with the spine surgeon – to be incorporated into the Dialogue Support algorithm. This way, the prediction values of each study participant would be readily available as patient and surgeon decide upon treatment.

Inclusion in the study began in the spring of 2022 and is set to be finished in 2025 as shown in Table 2. Follow-up is at 12 months after surgery or 12 months after the decision of non-surgical treatment.

Sample size

Today, approximately 65% of the patients with CR and 53% with LSS registered in Swespine report a successful outcome (i.e., responded "pain free" or "much improved" on GA arm pain and GA leg pain respectively) after surgery for degenerative spine conditions [6]. A 10% unit increase in successful outcome (i.e., a 75% and 63% success rate respectively) was considered an indication of the tool affecting the outcome. For the sample size calculations, we used standardized difference as provided by Altman and Fisher's exact test [21]. When power was set at 80% and the significance level at 0,05, the CR surgically treated group would require 348 patients and the LSS surgically treated group would require 401 patients. With an estimated drop-out at follow-up of approximately 20%, the size increased to 435 patients in the CR group and 500 in the LSS group, see Fig. 2.

Inclusion in the non-surgery group will end when 435 and 500 participants have been included in each corresponding surgically treated group. Power was not calculated on this group since the purpose for collecting these

Table 2 Study activities

	Staff member	Prestudy screening and consent	Study visit	Follow-up 1 year
Participant referred to spine clinic for assessment of Lumbal Spinal Stenosis or Cervical Radiculopathy	Referral reviewing surgeon	X		
MRI < 1 year old concurring with symptoms according to referral	Referral reviewing surgeon	X		
Written participant information and consent form sent to patient	Medical secretary	X		
Participant fills out baseline variables in waiting room before clinical assessment by surgeon			X	
Dialogue Support prediction of outcome is discussed as patient and surgeon decide upon treatment	Surgeon		X	
Follow-up form sent to participant	Swespine administration (for surgical treatment) Study administration (for non-surgical treatment)			X

Magnetic resonance imaging (MRI)

participants is to monitor any selection bias. The size of the group is estimated to be approximately 150.

The experimental groups will be compared to patients, registered in Swespine during 2013–2018 (i.e., before the introduction of the Dialogue Support), using multivariable regression methods as well as propensity score matching (see below). The use of these statistical methods is considered appropriate since the number of LSS patients reporting at the one-year follow-up has exceeded 5000 each of these years and the number of CR patients exceeded 500. Thus, the chance of finding suitable controls should be more than fair.

Recruitment

As stated above, all patients referred to the clinics for LSS or CR will be screened for inclusion. Further clinics may be approached and included if the targeted sample size shows to be difficult to collect within the time plan.

Data collection methods

Data on the experimental groups are collected prospectively and will be stored in the Swespine database, where a specific registration application has been created for the inclusion of participants (see Participant timeline). Data on the control groups will be collected retrospectively from Swespine.

Mortality in both the experimental, control and non-surgical groups will be acquired from the Swedish National Cause of Death Register [22].

A controller will ensure that the patients receive the correct follow-up. The controller will also see to that any crossovers receive the correct follow-up questionnaires.

Data management

All data will be collected electronically through the Swespine registration application created for the trial. The data will be stored on the Swespine servers which have a high level of security complying with necessary national laws for patient data [23].

Analysis populations

The analysis populations for the main analysis of the surgical group, as well as of the non-surgical group, will consist of full analysis set (FAS) individuals. Time until death will be performed on all participants.

Statistical methods

Data analysis will be performed after the entire sample has been collected. Analysis will be performed using SAS Software version 9.4 or later (SAS Institute Inc., Cary, NC, USA). The significance level is established at 0.05 and the confidence interval is set at 95%. The surgically as well as the non-surgically treated groups will be matched to control groups with similar patient characteristics on which the Dialogue support has not been used. Both propensity score-matching and a covariate-adjusted multivariate regression model will be performed. The adjusted analysis will be the primary analysis while the propensity-score-matching will be used as a supportive analysis to validate the strength of the analysis. Propensity-score matched analysis is expected to have lower power.

Baseline variables will be presented using mean (±standard deviation), or median and interquartile range or range as appropriate, or as counts (percentages). Baseline variables will be compared between the study population and the control group. For difference between two groups, ordered categorical variables will be analyzed using Mantel-Haenszel chi-square trend test, non-ordered by Pearson chi-square test and dichotomous by Fisher's exact test. Continuous variables will be compared between two groups using student t-test or Mann-Whitney U test depending on the variable distribution.

Covariate-adjusted multivariate model

A covariate-adjusted multivariate model will be applied to compare the surgery group with the control group. All baseline characteristics used in the Dialogue Support (see Table 1) will be used as covariates. The correlation between the baseline characteristics has previously been examined during the creation of the Dialogue Support [8]. The relative difference between the groups will then be calculated using a multivariable logistic regression model where the GA and Satisfaction will be set as dependent variables and the covariates as independent variables together with group variable as the main effect variable. C-statistics and Hosmer-Lemeshow test will be calculated for the model performance and goodness-of-fit, respectively.

Propensity score matching

A propensity score matching will also be used to balance the observed covariates simultaneously between the groups to improve the crude estimate of the effect of the Dialogue Support. The statistician will create a logistic regression studying probability for surgery using the Dialogue Support vs. not using the Dialogue Support based on patients' baseline characteristics (see Table 1) which will be used to derive the propensity score. As a matching algorithm a 1:1, nearest neighbor matching where each patient in the observed group will be compared to a unique patient in the control group based on the nearest propensity score. The optimal caliper width will be achieved by calculation 0.2 of the standard deviation of the logit of the propensity score [24]. To validate the model, the significance and standardized mean difference of propensity scores between the pre- and post-matched samples will be compared. Logistic regression analysis will then be performed to assess the effect on the primary outcomes of Global assessment and satisfaction. Hosmer-Lemeshow test and C-statistics will be calculated for the regression model to measure the goodness of fit and discriminative ability respectively [25].

Selection bias analysis

Analysis between the non-surgical group and a matched control group will be performed to study if any differences in results may depend on selection bias. Thus, changes in the EQ-5D between baseline and follow-up will be calculated. Both the index, as well as the EQ-VAS, and separate dimensions will be analyzed. Whether any statistically significant changes are also clinically relevant will be difficult to conclude in the absence of MIC estimates. However, it will be possible to make comparisons to the normal age-matched Swedish population, and the matter will be deliberated in the Discussion section. The selection bias analysis will be performed using linear models if the dimension has a normal distribution or Mann-Whitney U test if the dimension does not have a normal distribution [26].

A Cox regression analysis will be performed on time to one-year mortality between the non-surgically and surgically treated groups, adjusted for covariates that differ between the groups and that are significant predictors to the outcome, i.e., per definition being confounders. Assumption of proportional hazards will be checked. In case the assumption is not met, a method including the follow-up time in the model will be used, such as extended Poisson regression. Effect size will be presented by hazard ratios and their 95% CI.

Missing data

Age, and number of screening failures, i.e., patients who declined inclusion, will be noted at each participating clinic.

As study participants as well as spine surgeons fill out the digital study forms, all the required data must be imputed to proceed to the study version of the Dialogue Support. Therefore, no missing variables in the baseline should occur.

A controller will ensure that the surgeons fill out the surgery form and that it is registered in the database.

All participants will receive their follow-up questionnaire after one year. One reminder will be sent. If the participant does not answer the follow-up they will be excluded from the principal analysis, mortality analysis will be performed. Hence, multiple imputation methods will not be used.

A missing data analysis will be performed on patients that are lost to follow-up to study if their characteristics differ from the participants completing follow-up.

Discussion

Terms like AI, machine learning and big data are now imbuing everything around us and are revolutionizing the collection and processing of data in a way unimaginable only a few years ago. AI utilization can now be found in the production industry, the academia, the transportation sector and the financial and judicial systems [27] and there is no denying that it soon will be a prominent feature also in the healthcare system. ML applications are becoming increasingly important in decision making [27], and when employed within health care, warrant a high level of usability and interpretability to gain acceptance, as they are intended as an aid in potentially life changing decisions. However, to the layman, it is not intuitive how these systems work or how they reach their conclusions and there is a risk that an ML function will make decision-making and information collection harder to understand and grasp and also to trust, if the underlying algorithms are not sufficiently understandable [28]. The Dialogue Support has the prospective to make large

amounts of data interpretable for surgeons and individual patients, potentially enhancing the accuracy of the predicted outcome and making more informed decisions possible. By studying the effects of this already existing algorithm, an important step towards an implementation of ML within the medical field is taken and the results may be used to improve the development of these more advanced deep learning algorithms in a desired direction.

Many predictive models have been proposed within degenerative spine surgery and as most of these models are only internally validated their use is limited, since the impact on the outcome of surgery is unknown [1]. If the current trial demonstrates that patients to a larger degree have a better outcome when the Dialogue Support was used preoperatively, one can attribute this difference between study participants and the retrospective control group to the use of the algorithm, as the patient-reported postoperative outcome has been stable for the past 10 years in Swespine data for LSS and CR patients, i.e. there is a low risk of difference due to other factors such as improvements in surgical technique [29].

A prospective non-randomized trial design was selected over the gold standard randomized controlled trial (RCT) as the Dialogue Support has been publicly available since 2019 and may consequently be accessed by study participants before or during the trial, leading to potential bias. The advantage of an RCT would have been that the effect of the Dialogue Support could have been studied with less potential confounders and bias, had the participants not been able to access the Dialogue Support. This is a limitation to the current trial. Another limitation is the absence of MIC value for the EQ-5D, making it unclear whether statistically significant score change provides clinically relevant improvement for the individual patient.

Considering the limited resources within most health care systems, health care professionals and policy makers should strive to use all available tools to maximize the beneficial effect on the population. When predicted outcome is individualized and reliable, patients with high probability of improvement can proceed to surgery, whereas patients with comorbidity can refrain from spinal surgery if probability of postoperative improvement is low, thereby avoiding the risks associated with spinal surgery. In addition to positive effects for the individual patient, there may be a financially beneficial consequence on the health care system. This will be the largest external validation of a prediction model within degenerative spine surgery to date.

By conducting a multicenter study including both private and public hospitals in several different regions of Sweden, an indicative sample of the population will be collected, providing a strong validation of the Dialogue Support. A continuous inclusion of other healthcare providers from different parts of Sweden will be performed during the trial period.

When using a predictive algorithm there is a significant risk of selection bias, as the surgeon might be more prone to recommend non-surgical treatment if the predicted probability of successful outcome is poor. The current trial will examine this group of patients, and by comparing their quality of life and mortality with surgically treated patients, assess if there is indeed a selection bias when using a predictive algorithm [30].

The hype regarding Machine Learning is hard to miss. The question is if it can live up to our expectations. The term is broad and covers a multitude of applications. The Dialogue Support is based on a regression model, with an annual update of the database. A further development of the tool is intended, but first, its effects in clinical practice should be tested.

A natural course of development of the Dialogue Support would be an "upgrade" to a ML model. However, the potential influence of ML-based prediction models on the surgeon-patient shared decision-making process is not known and is according to a recently published scoping review yet to be studied [31]. An advantage of using a classical logistic regression model like the one used in the Dialogue support is that it may be easier to understand and explain to patients as compared to a ML prediction model, but this is an area that needs to be studied further [32].

In summary, we hope that this study of the Dialogue Support will provide evidence on whether the use of an advanced digital decision tool can lead to a better patient-reported outcome after surgery.

Abbreviations

ODI	Oswestry Disability Index
NRS	Numeric Rating Scales
ML	Machine Learning
CR	Cervical radiculopathy
LSS	Lumbar spinal stenosis
PROM	Patient-reported outcome measures
GA	Global Assessment
AUC	Area Under Curve
EQ-5D	Euroqol-5-Dimensions quality of life questionnaire
NDI	Neck Disability Index
MIC	Minimal Important Change
FAS	Full analysis set
RCT	Randomized controlled trial

Acknowledgements

The authors would like to acknowledge the participating spine units: Sahlgrenska university hospital, Gothenburg, Akademiska sjukhuset, Uppsala, Capio Spine Center Gothenburg, Capio Spine Center Stockholm, Ryggkirurgiskt Centrum Stockholm, Aleris Elisabethsjukhuset, Uppsala, Aleris Ortopedi Ängelholm, Aleris Ortopedi Malmö. The authors would also like to thank our statistician, Aldina Pivodic.

Author contributions

OH and CP designed the trial. Ethical applications were handled by CP. EE, LVA and CP have contributed to the contents of the manuscript in consultation with OH and PF . All authors have shared in reviewing and have given their

approval of the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding
The trial has received research grants from AFA Insurance, the Felix Neubergh foundation, and a stipend from the Society of Swedish Spine Surgeons.

Data availability
All data collected during the study will be available by the corresponding author on reasonable request. All documents will be stored in a safe online database, accessible only to the researchers involved. All names and personally identifiable information will be encoded. The statistician that will do the final analysis will receive the coded data.

Declarations

Ethics approval and consent to participate
The trial has been approved by the Swedish Ethical Review Authority (Approval No: 2021–03503). Participation in the trial has been approved by the director of department on each participating clinic. The trial dataset will be accessible only to the responsible researchers and a participating statistician.

Consent for publication
Not applicable.

Consent or assent
An informed consent form will be sent to the patients along with information on the study. When the patient arrives at the outpatient clinic, the staff will collect the signed form before adding the patient to the study database. These consent forms will be kept in a secure compartment.

Confidentiality
Information regarding the patients will be kept on the Swespine servers using all necessary security procedures. All participants will prior to publication be de-identified and all data will be analyzed and presented on a group level. The data will be kept for 10 years and thereafter deleted.

Dissemination policy
There will be no direct communication of the results to the patients. The results will be presented in scientific papers submitted to an international journal. The results will also be summarized in Swedish in the national journal for physicians, which is publicly available.

Competing interests
The authors declare no competing interests.

Author details
[1]Hallands Sjukhus Varberg, Varberg, Sweden
[2]Department of Surgery and Orthopaedics, Alingsås Lasarett, Alingsås, Sweden
[3]Capio Spine Center Göteborg, Gruvgatan 6, Västra Frölunda 421 30, Sweden
[4]Futurum Academy for Health and Care, Hus B4, Lanssjukhuset Ryhov, Jonkoping 553 05, Sweden
[5]RKC Spine Center, Sodra Fiskartorpsvägen 15H, 114 33, Stockholm, Sweden
[6]Department of Orthopedics, Sahlgrenska University Hospital, Gothenburg, Sweden
[7]The Sahlgrenska Academy, Gothenburg University, Gothenburg, Sweden

Received: 10 May 2023 / Accepted: 19 July 2024
Published online: 21 August 2024

References

1. Lubelski D, Hersh A, Azad TD, Ehresman J, Pennington Z, Lehner K, et al. Prediction models in degenerative spine surgery: a systematic review. Global Spine J. 2021;11(1_suppl):s79–88.
2. Stephens ME, O'Neal CM, Westrup AM, Muhammad FY, McKenzie DM, Fagg AH, et al. Utility of machine learning algorithms in degenerative cervical and lumbar spine disease: a systematic review. Neurosurg Rev. 2022;45(2):965–78.
3. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. New York, NY: Science; 2015;349(6245):255–60.
4. Pedersen CF, Andersen MO, Carreon LY, Eiskjaer S. Applied Machine Learning for Spine surgeons: Predicting Outcome for patients undergoing treatment for lumbar disc herniation using PRO Data. Global Spine J. 2022;12(5):866–76.
5. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. JAMA. 2018;319(13):1317–8.
6. Swespine. annual report 2022. https://www.swespine.se: Swedish Society of Spinal Surgeons; 2022.
7. Iderberg H, Willers C, Borgstrom F, Hedlund R, Hagg O, Moller H, et al. Predicting clinical outcome and length of sick leave after surgery for lumbar spinal stenosis in Sweden: a multi-register evaluation. Eur Spine Journal: Official Publication Eur Spine Soc Eur Spinal Deformity Soc Eur Sect Cerv Spine Res Soc. 2019;28(6):1423–32.
8. Fritzell P, Mesterton J, Hagg O. Prediction of outcome after spinal surgery-using the dialogue support based on the Swedish national quality register. European spine journal: official publication of the European Spine Society, the European Spinal Deformity Society, and the European section of the cervical. Spine Res Soc. 2022;31(4):889–900.
9. Parai C, Hagg O, Lind B, Brisby H. The value of patient global assessment in lumbar spine surgery: an evaluation based on more than 90,000 patients. Eur Spine Journal: Official Publication Eur Spine Soc Eur Spinal Deformity Soc Eur Sect Cerv Spine Res Soc. 2018;27(3):554–63.
10. Lurie J, Tomkins-Lane C. Management of lumbar spinal stenosis. BMJ (Clinical Res ed). 2016;352:h6234.
11. Iyer S, Kim HJ. Cervical radiculopathy. Curr Rev Musculoskelet Med. 2016;9(3):272–80.
12. Pedersen CF, Andersen MO, Carreon LY, Eiskjaer S. Validating the predictive precision of the dialogue support tool on Danish patient cohorts. N Am Spine Soc J. 2023;13:100188.
13. Hagg O, Fritzell P, Oden A, Nordwall A, Swedish Lumbar Spine Study G. Simplifying outcome measurement: evaluation of instruments for measuring outcome after fusion surgery for chronic low back pain. Spine. 2002;27(11):1213–22.
14. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine. 2000;25(22):2940–52. discussion 52.
15. Treadwell JR, Lucas S, Tsou AY. Surgical checklists: a systematic review of impacts and implementation. BMJ Qual Saf. 2014;23(4):299–318.
16. Burstrom K, Teni FS, Gerdtham UG, Leidl R, Helgesson G, Rolfson O, et al. Experience-based Swedish TTO and VAS Value Sets for EQ-5D-5L Health states. PharmacoEconomics. 2020;38(8):839–56.
17. Chiarotto A, Maxwell LJ, Ostelo RW, Boers M, Tugwell P, Terwee CB. Measurement Properties of Visual Analogue Scale, Numeric Rating Scale, and Pain Severity Subscale of the brief Pain Inventory in patients with Low Back Pain: a systematic review. J Pain. 2019;20(3):245–63.
18. Steinhaus ME, Iyer S, Lovecchio F, Khechen B, Stein D, Ross T, et al. Minimal clinically important difference and substantial clinical benefit using PROMIS CAT in cervical spine surgery. Clin Spine Surg. 2019;32(9):392–7.
19. Parai C, Hägg O, Lind B, Brisby H. ISSLS prize in clinical science 2020: the reliability and interpretability of score change in lumbar spine research. European spine journal: official publication of the European Spine Society, the European Spinal Deformity Society, and the European section of the cervical. Spine Res Soc. 2020;29(4):663–9.
20. Pool JJM, Ostelo RWJG, Hoving JL, Bouter LM, de Vet HCW. Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with Neck Pain. Spine. 2007;32(26):3047–51.
21. Altman DG. Practical statistics for medical research. 1st ed. London; New York: Chapman and Hall; 1991. xii, 611 p. p.
22. Kruse M, Thoreson O. The prevalence of diagnosed specific back pain in primary health care in Region Vastra Gotaland: a register study of 1.7 million inhabitants. Prim Health Care Res Dev. 2021;22:e37.
23. Stromqvist B, Fritzell P, Hagg O, Jonsson B. Swedish Society of Spinal S. The Swedish spine Register: development, design and utility. European spine journal: official publication of the European Spine Society, the European Spinal Deformity Society, and the European section of the cervical. Spine Res Soc. 2009;18(Suppl 3):294–304.
24. Huelsenbeck JP, Ane C, Larget B, Ronquist F. A bayesian perspective on a non-parsimonious parsimony model. Syst Biol. 2008;57(3):406–19.

25. Kim RB, Garcia RM, Smith ZA, Dahdaleh NS. Impact of Resident Participation on outcomes after single-level Anterior Cervical Diskectomy and Fusion: an analysis of 3265 patients from the American College of Surgeons National Surgical Quality Improvement Program Database. Spine (Phila Pa 1976). 2016;41(5):E289–96.
26. Nohr EA, Liew Z. How to investigate and adjust for selection bias in cohort studies. Acta Obstet Gynecol Scand. 2018;97(4):407–16.
27. Islam MR, Ahmed MU, Barua S, Begum S. A systematic review of Explainable Artificial Intelligence in terms of different application domains and tasks. Appl Sci. 2022;12(3):1353.
28. Combi C, Amico B, Bellazzi R, Holzinger A, Moore JH, Zitnik M, et al. A manifesto on explainability for artificial intelligence in medicine. Artif Intell Med. 2022;133:102423.
29. Cabitza F, Campagner A, Ronzio L, Cameli M, Mandoli GE, Pastore MC, et al. Rams, hounds and white boxes: investigating human-AI collaboration protocols in medical diagnosis. Artif Intell Med. 2023;138:102506.
30. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in Health Care services caused by the Adoption of Digital Health Technologies: scoping review. J Med Internet Res. 2022;24(3):e34144.
31. Toh ZA, Berg B, Han QYC, Hey HWD, Pikkarainen M, Grotle M, et al. Clinical decision support system used in spinal disorders: scoping review. J Med Internet Res. 2024;26:e53951.
32. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206–15.

Publisher's Note