Article

Machine learning algorithms provide greater prediction of response to SCS than lead screening trial: A predictive AI-based multicenter study

Amine Ounajim [1,2], **Maxime Billot** [1], **Lisa Goudman** [3,4], **Pierre-Yves Louis** [5,6], **Yousri Slaoui** [2], **Manuel Roulaud** [1], **Bénédicte Bouche** [1], **Philippe Page** [7], **Bertille Lorgeoux** [1], **Sandrine Baron** [1], **Nihel Adjali** [1], **Kévin Nivole** [1], **Nicolas Naiditch** [1,8], **Chantal Wood** [1], **Raphaël Rigoard** [9], **Romain David** [1,10], **Maarten Moens** [3,4] and **Philippe Rigoard** [1,7,11]

[1] PRISMATICS Lab (Predictive Research in Spine/Neuromodulation Management and Thoracic Innovation/Cardiac Surgery), Poitiers University Hospital, 86021 Poitiers, France.

[2] Laboratoire de Mathématiques et Applications, UMR 7348, Poitiers University and CNRS, 86000 Poitiers, France.

[3] Department of Neurosurgery, Universitair Ziekenhuis Brussel, 1090 Brussels, Belgium;

[4] STUMULUS research group, Vrije Universiteit Brussel, 1090 Brussels, Belgium;

[5] PAM UMR 02.102 Université Bourgogne Franche-Comté, AgroSup Dijon, 21000 Dijon, France

[6] Institut de Mathématiques de Bourgogne, UMR 5584 CNRS, Université Bourgogne Franche-Comté, 21000 Dijon, France.

[7] Department of Spine Surgery & Neuromodulation, Poitiers University Hospital, 86021 Poitiers, France.

[8] Dyname, UMR 7367, Faculty of Social Sciences, University of Strasbourg, 67083 Strasbourg, France;

[9] Data architect, CEA Cadarache, Service des Technologies de L'Information et de la Communication, Ingénierie des Systèmes d'Information, 13115 Saint-Paul Lez Durance.

[10] Physical and Rehabilitation Medicine Unit, Poitiers University Hospital, University of Poitiers, 86021 Poitiers, France.

[11] Pprime Institute UPR 3346, CNRS, ISAE-ENSMA, University of Poitiers, 86360 Chasseneuil-du-Poitou, France.

* Correspondence: amine.ounajim@chu-poitiers.fr; Tel.: +33752415229;

Abstract Persistent Pain after Spinal Surgery can be successfully addressed by Spinal Cord Stimulation (SCS). International guidelines strongly recommend that a lead trial be performed before any permanent implantation. Recent clinical data highlight some major limitations of this approach. First, it appears that patient outcomes, WITH OR WITHOUT lead trial, are similar. In contrast, during trialing, infection rate drops drastically within time and can compromise the therapy. Using composite pain assessment experience and previous research, we hypothesized that machine learning models could be robust screening tools and reliable predictors of long-term SCS efficacy. We developed several algorithms including logistic regression, Regularized Logistic Regression (RLR), naive Bayes classifier, artificial neural networks, random forest and gradient boosted trees to test this hypothesis and to perform internal and external validations, the objective being to confront model predictions with lead trial results using a 1-year composite outcome from 103 patients. While almost all models have demonstrated superiority on lead trialing, the RLR model appears to represent the best compromise between complexity and interpretability in prediction of SCS efficacy. These results underscore the need to use AI based-predictive medicine, as a synergistic mathematical approach, aimed at helping implanters to optimize their clinical choices on daily practice.

Keywords spinal cord stimulation; screening trial; lead trial; infection; supervised learning; machine learning; predictive modeling; patient outcome.

1. Introduction

Failed back surgery syndrome (FBSS), now known as Persistent Spinal Pain Syndrome type 2 (PSPS-T2), is characterized by persisting back and/or leg pain despite one or several spinal surgical procedures [1–4]. PSPS-T2 incidence remains devastating, affecting 10 to 40% of operated spine patients [5,6]. This generates severe social [7], financial and psychological burdens for a significant number of patients [8]. Given this context, PSPS-T2 patients are referred to a large panel of therapies through multidisciplinary team pain management and when refractory, they can be successfully treated with Spinal Cord Stimulation (SCS) [9–14]. SCS outcomes rely on patient selection, which remains challenging since implanters have to face the nature of pain, infiltrating all dimensions of patient quality of life [15], defining different trajectories for different patient profiles, and impacting the most vulnerable refractory PSPS patients, potentially eligible to SCS, with an extreme variety of clinical presentations.

With a legitimate determination to bring some medico-economic rationale to technological evolution, health care systems provide strict rules of SCS implantation since its initial diffusion; following international recommendations, a lead trial must be performed and validated before any permanent device implantation [16–19]. In addition, SCS lead trialing is intended to potentially optimize neural structure spatial targeting, as some implanters choose to perform intra-operative table testing and/or deliberate temporary trial to optimize lead placement, and to reimplant the same patient later permanently, when test results are positive [20]. Moreover, from a patient's perspective, lead trialing offers an opportunity to "mimic" the potential added value of SCS during an average 5-10 day-period of trial [21].

However, recent clinical data highlight some major limitations of this approach:

-   First, one can observe a lack of homogeneity among practices, making comparisons and recommendations regarding lead trialing very difficult: a patient can be trialed with a surgical lead (requiring an invasive approach) vs a percutaneous trial, where one, two or three percutaneous lead(s) are placed in the spinal canal, under local sedation or general anesthesia [14,18]; he can be trialed from 3 to 28 days, depending on healthcare systems [22], which is responsible for significant differences in terms of lead trial complication rates [14]; during the trial, patients can test one, a few, or hundreds of programs [18], which might affect lead trial outcome, depending on the clinical expertise of the implanting team and trial duration. Ultimately, lead trialing could create a bias since the trial cannot be blinded except in paresthesia-free techniques at the price of strict protocols, which cannot be standardized among practices, centers and countries, except for research purposes.

-   Lead trial outcomes have been reported using Visual Analogue Scale (VAS) or Numeric Pain Rating Scale (NPRS). Based on these assessments, guidelines recommend that a refractory PSPS patient can be eligible to permanent implant, if a 30-50% of pain decrease is observed during the trial [14,16,17,19]. It appears that this unidimensional modality of pain assessment can no longer be considered as the only gold standard to delineate the implanting physician's guidelines and international recommendations, since this would reflect only one dimension of the patient's quality of life, needs and expectations [23]. While composite multidimensional pain indexes, following the application of pain therapy on chronic refractory patients [23], would help to capture the essence of pain substrate and pain potential relief, they are not currently part of the recommended pain assessment toolbox.

-   Third, it appears that patient selection, final implantation ratio and patient outcomes, WITH OR WITHOUT lead trial, are similar [24]. Indeed, several studies have evaluated the ability of screening trial to predict the long-term efficacy of SCS [24–26]. In a multicentric randomized controlled trial, Eldabe et al. [24] compared pain relief at 6-month follow-up between 54 patients who underwent a screening trial and 51 patients who did not. They found no significant difference regarding pain relief between groups. Furthermore, they reported that the diagnostic accuracy of the screening trial presented sensitivity of 100% (percentage of patients with a positive screening trial

among responders) and specificity of 17% (percentage of patients with a negative screening trial among non-responders). All in all, the authors concluded that outcomes do not differ with or without screening trial, and that screening trial was not able to identify long-term non-responders.

- Fourth, it has been shown that infection rates during SCS trials increase logarithmically within time, after 14 days of trialing, and can compromise the therapy [14].

To date, we can conclude there is no consensus on the usefulness of the screening trial phase due to the invasiveness of the procedure and potential complications such as lead migration, dural puncture during lead placement and/or infection [27,28]. Aiming to find the best compromise between risks and benefits, several authors have attempted to find alternatives to screening trial, studying factors such as age [29,30], sex [31,32], psychological state [30,33–35] and pain duration [36,37], their objective being to demonstrate that they are associated with long-term pain relief following SCS. However, only classical statistical methods including logistic regression models and correlation analyses have been applied to identify these predictors, which renders their transposability to daily practice quite difficult, if not impossible. Recently, Goudman *et al* showed that machine learning models can achieve good performance for predicting the efficacy of high-frequency SCS in patients with PSPS-T2 [38], which is highly specific and only applicable to high-frequency SCS; this represented a first step at use of predictive medicine in this field.

Using our experience of multidimensional composite pain assessment and previous predictive research [15,23], we hypothesized that machine learning models, based on homogeneous data pooled from several multicenter studies, could be used as robust screening tools to predict SCS efficacy. To test this hypothesis, we developed, for the purpose of this study, several algorithms designed to confront model predictions and to compare our model performances to lead trial results among 103 implanted patients.

2. Materials and Methods

2.1 Patient data

Data from two different prospective comparative studies were used to conduct this work.

First dataset

The first study is ESTIMET [18], which is a multicenter randomized controlled trial, including 115 PSPS-T2 patients eligible for SCS and implanted with surgical multicolumn SCS paddle lead, in 12 French centers with a 1-year follow-up. The study details are available at https://clinicaltrials.gov/ct2/show/NCT01628237. The primary objective of this study was to compare the efficacy of multicolumn SCS programming to the efficacy of monocolumn SCS programming. As part of the ESTIMET study, all subjects provided informed consent and enrolled following ethical committee approval (CPP-Ouest III) [18]. The study population consisted of PSPS-T2 patients with refractory pain, eligible to SCS according to the French guidelines for SCS selection and implantation. Per these guidelines, an average of 7-day screening trial period was mandatory for all study patients. Patients with a 50% pain decrease, or patients for whom the improvement was clinically important according to a patient-implanter agreement, were implanted with a permanent SCS device at the end of the trial. Among the ESTIMET study patients, those who did not try Transcutaneous Electrical Nerve Stimulation (TENS) were removed from the analysis because TENS efficacy belongs to the set of predictive variables used in the development of the models in this paper (Figure 1). Finally, ninety-one patients who underwent TENS therapy, completed baseline data, and completed the study 1-year follow-up were included in the analysis. ESTIMET study data was used for the training and internal validation process of our models.

Second dataset

The second study is AIVOC (https://clinicaltrials.gov/ct2/show/NCT02821897), which is a monocentric comparative study, including 15 patients implanted with SCS under general vs awake anesthesia at Poitiers University Hospital (France), with a 1-year follow-up. This study examines the effect of target-controlled intravenous infusion on SCS implantation, lead placement optimization using patient intra-operative feedback and SCS efficacy on back pain coverage. Patients in this study were randomized to either be implanted using general anesthesia during lead implantation or to be implanted using target-controlled intravenous anesthesia with active patient-implanter cooperation during the surgery. Three patients were lost to follow-up. The 12 remaining patients who underwent 12-month follow-up from the AIVOC study were used for external validation.

For both the ESTIMET and the AIVOC studies, following verification of inclusion/non-inclusion criteria patients were included and evaluated at baseline, in terms of their sociodemographic, psychological, radiological, and clinical characteristics. One-month after inclusion, all patients were implanted with SCS and underwent a permanent trial: An average of 7-day screening trial period was mandatory for all, per French recommendations of ministry of Health. Patients with a 50% pain decrease, or patients for whom the improvement was clinically important according to a patient-implanter agreement, were implanted with a permanent SCS device at the end of the trial [19].
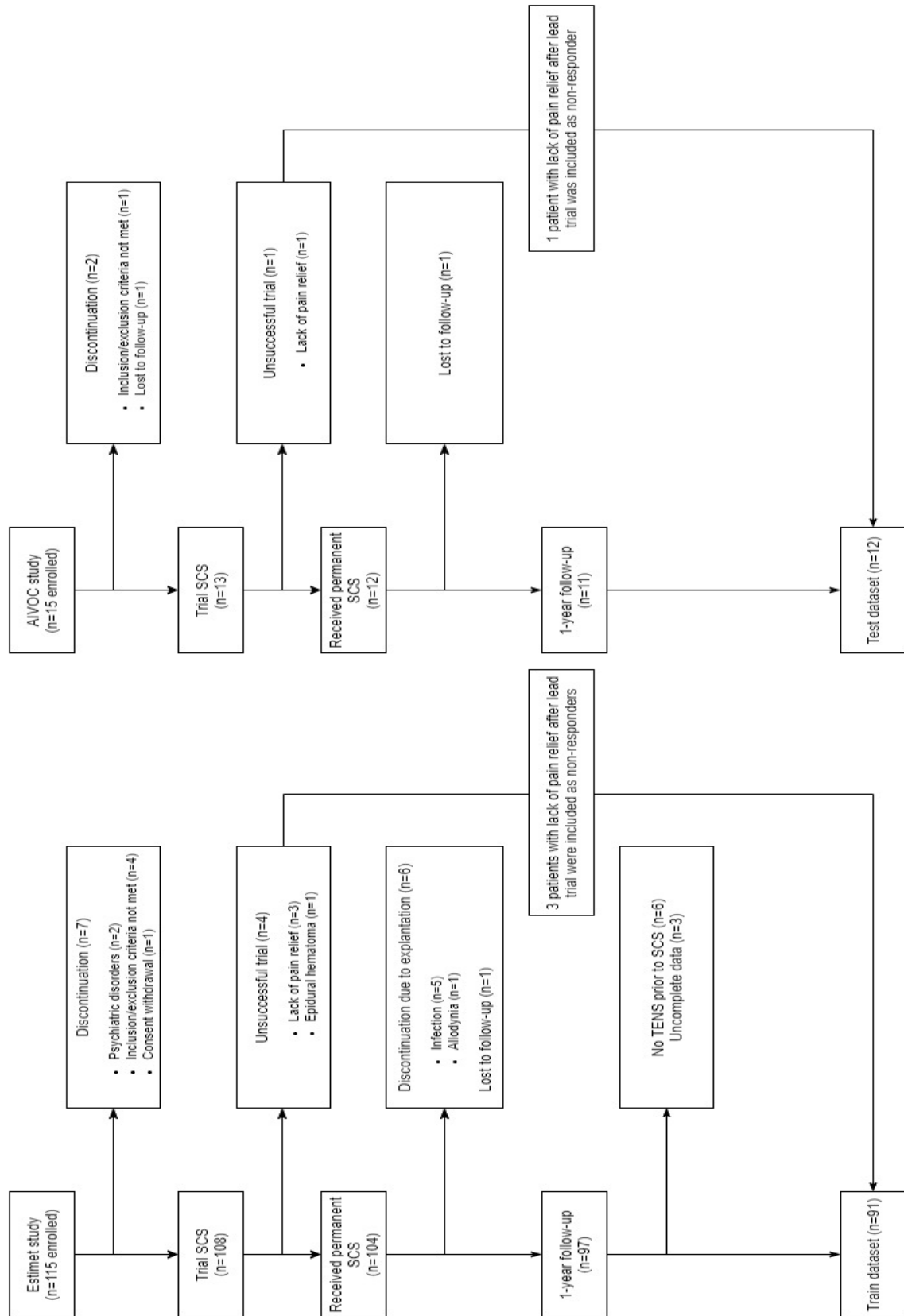
Figure 1. Flowchart of the patients who participated in ESTIMET and AIVOC included in this study.

2.2 Studied variables

2.2.1 Primary outcome

We evaluated SCS efficacy using pain intensity assessed by means of a Visual Ana-logic Scale (VAS), health-related quality of life (EuroQol with five dimensions and three levels (EQ5D-3L)) [39], functional disability evaluated using the Oswestry Disability Index (ODI) [40] and severity of depression using the Montgomery and Asberg Depression Rating Scale (MADRS) [41]. In order to achieve holistic evaluation, we used Principal Component Analysis (PCA) with one principal component, including the percentage of global VAS decrease (i.e. (VAS baseline – VAS 12-month)/VAS at baseline), percentage of EQ-5D increase, percentage of ODI decrease, percentage of MADRS decrease, between baseline and 12-month follow-up. The first principal component of the PCA was taken as a standardized Global Health Improvement Score (GHIS). Patients were considered as responders if they had a GHIS >= 0. Patients who had a negative screening trial or a GHIS < 0 were considered as non-responders. This outcome was used as a binary dependent variable in our SCS efficacy classification problem. We also evaluated the relationship be-tween this composite outcome and the standard outcomes used in SCS literature and rec-ommended in France by health authorities, which are 50% pain VAS decrease, 30% ODI decrease and 0.19-point EQ-5D index difference.

2.2.2 Predictors

To avoid any bias induced by variable selection based on statistical significance, no primary variable selection was conducted. We used the 14 variables that were the most widely studied in the SCS literature. There included age [29,30], sex [31], Depression score [30,34] measured using MADRS score, Body Mass Index (BMI) [42], pain syndromes as-sociated with nervous or somatic lesions (hypoesthesia, brush allodynia) [43], pain in-crease by movement or by sustained position, TENS efficacy [44], baseline EQ-5D index, baseline back and leg VAS, baseline ODI score, pain duration (in years) [36,37] and the Medication Quantification Scale (MQS III) to measure medication consumption in chronic pain [45].

2.3 Statistical methods

The statistical analyses were performed using R 3.6.0 software (R Foundation for Sta-tistical Computing, Vienna, Austria).

2.3.1 Descriptive analysis

Categorical variables were described by numbers and percentages, while quantita-tive variables were described by their means and standard deviations (SD) or by their median and interquartile range (IQR) depending on the skewness of the variable. No missing data imputation was performed.

2.3.2 Multivariate analysis

For this analysis, we developed several algorithms based on Logistic Regression (LR), Regularized Logistic Regression (RLR), Naive Bayes (NB) classifier, Artificial Neural Networks (ANN), Support Vector Machines (SVM), Classification And Regression Trees (CART), Random Forest (RF) and Gradient Boosted Trees (GBT) to test our hypothesis. Commonly used model performance metrics ROC curve and area under ROC curve (AUC) were used to evaluate model accuracy. ESTIMET study data (n=91) was used as a training set and AIVOC study data (n=12) was used as an independent testing set to con-front model predictions and to compare model performances to the 5 to 10-day trial re-sults. Leave-One-Out cross-validation and Monte Carlo cross-validation were used to as-sess internal validity and performance variability.

All data were standardized in order to facilitate interpretation and convergence of the models. Standardization was conducted by subtracting the mean and dividing by the standard deviation.

In this section, each model and its implementation are described briefly. The following classification models were used in this paper to predict SCS outcome at 12-month follow-up:

LR: This model was developed using the *glm* function available in R stats package. Even if logistic regression can only detect linear relations between variables, it is still widely used because of its simplicity and interpretability, and it has shown better performance on simple classification problems where classes can be linearly separated. Backward-forward (bidirectional) stepwise variable selection procedure was used in order to identify the best LR model based on the Akaike Information Criterion (AIC). To avoid overfitting, no interaction terms were included in the model.

RLR: This regularized generalized linear model model [46] was developed using the *glmnet* package. Regularization is a technique used to shrink or reduce insignificant effects in the logistic regression to zero. This technique enables the model to avoid overfitting because it reduces model variance. As our data contain few variables, we opted for the use of ridge regularization.

SVM: Support vector machine models are known for their classification capability, since SVM algorithms are computationally stable, and generalize well, giving a sufficient number of training examples [47]. Besides linear relations, SVM can detect nonlinearities by transforming input data into a space of higher dimensionality using a kernel function. For the purposes of this study, we used the radial basis function kernel to allow non-linearity. SVM was developed using the *svm* function available in the *e1071* R package [48]. The optimal cost and gamma parameters were identified through cross-validation.

NB: This classifier [49] was developed with default hyperparameters using the Naive Bayes function available in the *e1071* R package.

ANN: The neural networks model [50] was developed using the *keras* package [51], which is a popular deep learning Python package that has been added recently to R software as a package available in the Comprehensive R Archive Network (CRAN). To avoid overfitting, we used a relatively small ANN. The ANN model contained two hidden layers, each comprising eight nodes. To ensure model convergence, we trained the ANN model on standardized data (0 mean and unit variance). Sigmoid activation functions were used to allow non-linearity. The weights were estimated using the backpropagation method, which is a gradient-based optimization method. This allows error estimation at the output of the hidden layer neurons; thereby enabling the update of weights in the hidden layers by means of error gradients.

CART: The classification tree model was developed using the *rpart* function available in the package with the same name [52]. In a CART model, all patients initially belong to a simple node representing good and non-responder rates. Afterwards, the node is split, creating two new child nodes. The splitting is performed by choosing the predictor and the optimal split point (e.g. age>45) that differentiates good responders from non-responders. The algorithm stops when the observations inside the nodes are homogeneous in terms of the outcome variable and further splits are undesirable. The minimum number of observations in a node was set at five.

RF: The random forest [53] was trained using the *randomForest* function available in the R package with the same name [54].

GBT: This gradient boosting based model [55] was developed using the *xgboost* package [56].

Both RF and GBT are tree-based ensemble models, which are a type of models based on the principle that averaging the predictions of several small models could help to obtain a better model. Each of the two models uses a different ensemble learning technique. RF uses Bagging (Bootstrap aggregating), which can be described as follows: for each iteration, a decision tree model is created using data from a bootstrap sample drawn from the training set, independently of other iterations. After growing all the trees, each tree casts a unit vote for the outcome (good or non-responder) of a new observation. The final prediction for this observation is the average of predictions obtained from all the trees.

GBT, as the name suggests, uses a technique called boosting, whereby weighted combinations of decision trees are constructed into a stronger classifier in an iterative way (contrary to Random Forest, where the weights are uniform, and the trees are grown independently). The strongest classification tree is weighted to count more substantially in

the prediction of outcome. A tree that most accurately classifies examples that were mis-classified by the first tree is grown next. This procedure allows the trees that are weak on some examples to be compensated by another tree, which performs better on the same examples.

2.4 Testing data and model assessment

Cross-validation was used to identify optimal hyperparameters of the models and to assess their internal validity. Our cross-validation procedure goes as follows: We divide our training dataset into 10 separate datasets (10-fold cross-validation). One subset is kept for model assessment. 9-fold cross-validation is conducted on the remaining nine subsets to identify the optimal hyperparameters, which are then used to develop the models. The final models are then tested on the 10th subset. This process is conducted for all of the 10 folds. This procedure allows us to reduce evaluation bias associated with identifying optimal hyperparameters and evaluating the models on the same subset.

The models and screening-trial efficacy (percentage of pain decrease) were evaluated using sensitivity, specificity, accuracy, and Area Under ROC Curve (AUC). The cross validation means of these evaluation measures and their standard deviations are reported.

2.5 External validation

An independent data set of the 12 remaining patients from the AIVOC study was used for model assessment and external validation. Similarly, models were evaluated externally using sensitivity, specificity, accuracy, and Area Under ROC Curve (AUC).

3. Results

3.1 Descriptive analysis

Descriptive statistics of our training and testing data can be found in Table 1. The majority of our predictors were homogeneous between the training and testing datasets. 49.5% of the training dataset were male and 50.5% were female while 41.7% of the testing dataset were male and 58.3% were female. The mean age of our training sample was 47.7 (SD = 9.5) and the mean age of the testing sample was 49.5 (SD = 14.7). We observed some differences between the training and testing datasets. Patients in the training dataset were less likely to respond to TENS therapy than patients in the testing dataset (52.7% for training dataset vs 83.3% for testing dataset). Pain medication consumption was also higher in the training dataset (MQS of 24.5 (SD = 14.7) vs 5.6 (SD = 7.8) for the testing dataset.

The results of this PCA leading to the development of our composite outcome can be found in Table 2. The first component of our PCA explained 50.1% of the total variance. The ODI percentage of decrease had the highest weight in the first component (0.86) followed by VAS (0.81), MADRS (0.59) and EQ-5D (0.51). In our training dataset, 45 patients (49.5%) had a positive holistic response to SCS and 46 (50.5%) had a negative holistic outcome. Similarly, 6 patients (50%) had a positive outcome and 6 patients (50%) had a negative outcome in the testing dataset. Table 3 shows the relationship between our composite outcome and the classical pain assessment outcomes used in the literature.

Table 1. Descriptive statistics of our 12-month outcome and baseline predictive variables for both the training and testing datasets.

Variable	Train set descriptive statistics	Test set descriptive statistics
Response variable		
Good composite outcome	45 (49.5%)	6 (50.0%)
Bad composite outcome	46 (50.5%)	6 (50.0%)
Predictors at baseline		
Age	47.7 (9.5)	49.5 (14.7)

Sex		
Male	45 (49.5%)	5 (41.7%)
Female	46 (50.5%)	7 (58.3%)
BMI	27.4 (5.04)	24.6 (3.9)
Pain duration	12.2 (10.7)	15.8 (15.1)
ODI	50.5 (9.1)	44.7 (12.8)
MADRS	16.9 (10.4)	11.3 (8.3)
EQ-5D	0.38 (0.20)	0.54 (0.19)
EQ-5D VAS	45.8 (17.3)	51.1 (20.0)
Leg VAS	75.0 (11.3)	72.9 (16.0)
Back VAS	71.2 (15.1)	67.6 (21.8)
TENS efficacy		
Effective	48 (52.7%)	10 (83.3%)
Not effective	43 (47.3%)	2 (16.7%)
Hypoesthesia		
Yes	28 (30.8%)	2 (16.7%)
No	63 (69.2%)	10 (83.3%)
Allodynia		
Yes	22 (24.2%)	5 (41.7%)
No	69 (75.8%)	7 (58.3%)
Positional pain changes		
Yes	74 (81.3%)	9 (75.0%)
No	17 (18.7%)	3 (25%)
MQS	24.5 (14.7)	5.6 (7.8)

BMI: Body Mass Index, ODI: Oswestry Disability Index, MADRS: Montgomery-Asberg Depression Rating Scale, EQ-5D: EuroQol-5 Dimensions, VAS: Visual Analogic Scale, TENS: Transcutaneous Electrical Nerve Stimulation, MQS: Medication Quantification Scale.

Table 2. Composition of the first principal component of the PCA of our outcomes.

Variables changes (%) between baseline and 12 months	1st principal component loadings (50.1% of the total variance)
ODI	0.86
VAS	0.81
MADRS	0.59
EQ-5D score	0.51

ODI: Oswestry Disability Index, MADRS: Montgomery-Asberg Depression Rating Scale, EQ-5D: EuroQol-5 Dimensions, VAS: Visual Analogic Scale.

3.2 Training data results (internal validation)

We developed our models using the eight different binary classification methods described in the statistical methods section. The SVM model showed the highest performance metrics according to our cross-validation procedure results (AUC = 0.801; SD = 0.202). It had specificity of 81.3% (SD = 14.8%) and sensitivity of 80.7% (SD = 20.1%). The GBT model also showed good performances with lower variability between folds, which indicates a more stable model (AUC = 0.790 (SD = 0.105); Specificity = 80.0% (SD = 3.6%); Sensitivity = 70.1% (SD = 16.7%)). The two logistic regression models LR and RLR, showed results comparable to the previous models. The regularized logistic regression model had AUC of 0.781 (SD = 0.120), sensitivity of 69.8% (SD = 12.0%) and specificity of 73.0% (SD = 12.9%). Our logistic regression model had AUC of 0.779 (SD = 0.114), sensitivity of 70.9% (SD = 18.0%) and specificity of 72.8% (SD = 13.3%). The AUCs of the RF, naive Bayes and CART were 0.755 (SD = 0.123), 0.697 (SD = 0.153) and 0.657 (SD = 0.136) respectively. According to these results, we would recommend GBT, SVM or RLR models. The advantage of the RLR model is that it can be interpreted as a simple logistic regression model. AUC of the screening-trial was 0.670 with sensitivity of 79.5% and specificity of 52.4%.

Table 3. Relationship between GHIS outcome and VAS decrease, ODI decrease and improvement in EQ-5D.

Outcomes	Good composite outcome (GHIS >= 0)	Bad composite outcome GHIS < 0
50% global VAS decrease		
Yes	43 (93.5%)	8
No	3	37 (82.2%)
30% ODI decrease		
Yes	34 (73.9%)	8
No	12	37 (82.2%)
0.19 points change in EQ-5D		
Yes	30 (65.2%)	17
No	16	28 (62.2%)

EQ-5D: EuroQol-5 Dimensions; ODI: Oswestry Disability Index; GHIS: Global Health Improvement score; VAS: Visual Analogue Scale.

3.3 External validation

The results of our models using the external validation set can be found in Table 4. The results obtained on the external validation set were similar to those on the external set. The best performances were achieved using the RF model, the GBT model and the RLR model. While the RF model showed lower performance on the training set, the GBT and RLR model had good results on both the training and testing sets. Similarly to the training set, the testing set screening-trial results showed good sensitivity (100%) but bad specificity (33.3%). 40% of patients with a bad 1-year outcome had a 50% pain decrease following screening-trial.

Table 4. AUC, specificity and sensitivity of screening-trial and our model on the external validation set.

Model	True good outcome	True bad outcome
Screening-trial (AUC = 0.69)		
Good outcome	6 (Sensitivity = 100%)	4
Bad outcome	0	2 (Specificity = 33.3%)
LR (AUC = 0.72)		
Predicted good outcome	5 (Sensitivity = 83.3%)	2
Predicted bad outcome	1	4 (Specificity = 66.7%)
RLR (AUC = 0.81)		
Predicted good outcome	5 (Sensitivity = 83.3%)	2
Predicted bad outcome	1	4 (Specificity = 66.7%)
SVM (AUC = 0.75)		
Predicted good outcome	6 (Sensitivity = 100%)	2
Predicted bad outcome	0	4 (Specificity = 66.7%)
NB (AUC = 0.81)		
Predicted good outcome	5 (Sensitivity = 83.3%)	1
Predicted bad outcome	1	5 (Specificity = 83.3%)
ANN (AUC = 0.72)		
Predicted good outcome	5 (Sensitivity = 83.3%)	2
Predicted bad outcome	1	4 (Specificity = 66.7%)
CART (AUC = 0.72)		
Predicted good outcome	4 (Sensitivity = 66.7%)	1
Predicted bad outcome	2	5 (Specificity = 83.3%)
RF (AUC = 0.83)		
Predicted good outcome	5 (Sensitivity = 83.3%)	1
Predicted bad outcome	1	5 (Specificity = 83.3%)
GBT (AUC = 0.81)		
Predicted good outcome	5 (Sensitivity = 83.3%)	1
Predicted bad outcome	1	5 (Specificity = 83.3%)

3.4 Model interpretability

The majority of the models discussed and analyzed in this paper are black box models, meaning that it is difficult to extract useful information on how the variables interact

with SCS outcomes. However, this interpretability is sometimes disregarded in order to achieve more complexity. In this section, we will show the role of explanatory variables in decision-making for the logistic regression model. Table 5 shows the unstandardized coefficients, standardized coefficients and their 95% confidence intervals, for each variable.

Based on the logistic regression model results (Table 5), the baseline depression score had a significant effect on the outcome of the treatment (MADRS odds ratio = 0.908; p-value = 0.002). Depressive patients have a reduced chance of having a good outcome 1 year after the implantation of SCS device. Interestingly, we observed a greater probability of achieving a successful outcome not only in patients with hypoesthesia related to back pain (odds ratio = 10.59; p = 0.0008) but also in those with positional back pain symptoms (odds ratio = 4.48; p = 0.043). Patients who achieved a 50% pain decrease after TENS therapy before SCS had a better chance of successful SCS therapy at 1-year follow-up (odds ratio = 0.2691544; p value=0.016).

Table 5. Standardized coefficients of selected variables, confidence intervals and significance levels.

Variables	Unstandardized coefficients (β)	Standardized coefficients	95% CI	p-value
Intercept	-3.044	-0.070	[-0.586;0.447]	0.792
Duration of pain	-0.038	-0.041	[-0.939;0.116]	0.137
MADRS	-0.097	-1.012	[-1.653;-0.371]	0.002**
EQ5D VAS	0.032	0.554	[0.015;1.093]	0.044*
Leg VAS	0.040	0.449	[-0.09;0.988]	0.102
Hypoesthesia : Yes	2.361	1.096	[0.455;1.737]	0.0008***
TENS : Not effective	-1.312	-0.659	[-1.196;-0.122]	0.016*
MQS	-0.024	-0.449	[-1.061;0.163]	0.151
Positional pain changes : Yes	1.500	0.588	[0.017;1.159]	0.043*

MADRS: Montgomery-Asberg Depression Rating Scale, EQ-5D: EuroQol-5 Dimensions, VAS: Visual Analogic Scale, TENS: Transcutaneous Electrical Nerve Stimulation, MQS: Medication Quantification Scale.

4. Discussion

The goal of this paper was to investigate the comparative performance of machine learning models vs recommended lead screening trial to predict SCS efficacy. Our results show that machine learning techniques offer the opportunity to predict patient SCS long-term outcomes with higher accuracy than SCS trialing (when considering a failed SCS trial as a negative long-term response). Compared to previous research works aiming to identify SCS outcome predictors, the originality of this mathematical approach is based on computing these predictors in a multivariate statistical model in order to transpose this predictive model to PSPS-T2 patient real-life conditions [45]. This study also led us to test our multi-dimensional pain assessment approach, reflected in a composite index, based on principal component analysis conducted on pain intensity, functional disability, quality of life and psychological distress measures. We have shown that this holistic pain evaluation is significantly associated with these four pain dimensions.

4.1. Potential added value of a predictive model vs lead trialing

Applying invasive therapeutic strategies to vulnerable patients implies empiric choices for the implanting physician. To develop a strong therapeutic alliance, when convinced, the physician also has to convince his patient that the benefits/risk ratio is in favor of applying a new technique. In that sense, a lead screening trial appears very comfortable,

not only for the patient but also for the physician, as described with humor by an internationally renowned implanter: "The beauty of SCS is that you try it before you buy it…!".

Conversely, it is now well-documented that lead trialing is associated with substantial complications [24,26], including infection [57], which can turn from what one patient has described as a "honeymoon" into a nightmare, leading to SCS explantation (11 out of 108 patients implanted with a multicolumn lead in ESTIMET study [18]).

Reconsidering back the benefits/risks balance, one should examine the benefits of lead trialing carefully, if risks are considered as substantial. Eldabe *et al*'s recent study leads to implanter disappointment, by showing that SCS outcomes remain similar, whether using lead trialing to select patients for permanent implantation, or not [24]. Moreover, in a recent RCT, Thomson and al. demonstrated that patient preference would be in favor of a single-step approach.

Therefore, an original mathematical predictive approach could aspire to replace SCS trial procedure by a more objective, non-invasive and accurate strategy based on AI, using machine learning models to reduce the risks of complications associated with invasive surgical procedures by more efficiently selecting the patients who may benefit from spinal cord stimulation.

4.2. Machine learning model accuracy to predict SCS efficacy

We more accurately predicted the efficacy of SCS than the screening-trial (AUC=0.69 for screening-trial vs AUC=0.81 for the RLR model). As a starting point, easy to transpose in daily practice, we recommend the use of RLR model as it achieved a remarkable performance for both internal and external validation. In contrast with the NB, RF, and GBT showing even greater performance, the RLR model is also easily interpretable (It can be interpreted similarly to a logistic regression model) and variable selection is included in the model estimation procedure, which can simplify the creation of an automated decision tool.

Few authors have proposed models predicting SCS efficacy. The most recent and relevant paper was published by Goudman *et al* [45]. In their paper, they developed a logistic regression model for predicting high-dose SCS efficacy using data from 92 FBSS patients and a set of variables including age, sex, back and leg pain intensity, MQS, ODI, Pittsburgh sleep quality index, EQ-5D and second-order interactions between these variables. They achieved 90% specificity and sensitivity on an out-of-sample dataset consisting of 20% of their dataset. The model contained some very large coefficients compared to the scale of the variables. This might be due to the inclusion of a large number of interaction terms for an intermediate population sample size. Cross-validation results and screening-trial predictive value were not reported in this paper despite very good performances to detect high-frequency SCS responders. Sparkes *et al* [35] also proposed a logistic regression model for predicting pain decrease at 12-month of follow-up using data from 56 FBSS patients. Their model included age, sex, duration of pain, anxiety and depression scores, and coping strategies. These results need however to be taken with caution, due to the lack of out-of-sample validation.

Using a different strategy, Thomson *et al* [21] proposed a SCS decision tool, based on consensus recommendations from a panel of experts. This concept should be considered as a complementary approach, reflecting key opinion leaders' views, from a physician's perspective, as discussed above, and not as an opposed vision. One approach is based on perception and convictions; "Artistic Science", which characterizes the patient-physician relationship, is designed to facilitate best therapeutic decision in a non-empiric world. The other approach is based on mathematical accuracy; "Art of Science" should help physicians to reinforce their choices and convictions, based on robust and quantitative substrate. While SCS literature on outcome predictors is currently heterogeneous, data/evidence-based medicine is clearly needed to establish new insights in the SCS community.

4.3. SCS predictors of lead trial success and SCS long term-outcomes

Our logistic regression model shows that patients with higher levels of depressive symptoms are less likely to benefit from SCS whereas patients with a higher perceived health-related quality of life (EQ-5D) are more likely to achieve a good outcome following SCS. This corroborates our findings from a previous work identifying two distinct patient profiles [15], for which depression was associated with lower level of activity, resulting in lower quality of life these profiles are considered as pejorative SCS outcome predictors in a recent sociological paper focusing on socio-professional status of PSPS patients [58].

TENS efficacy was also significantly associated with SCS efficacy. This has been shown since 2011 in a prospective predictive study published by Mathew *et al*.

Lastly, back pain hypoesthesia and changes in pain depending on patient position are associated with a greater likelihood of a good outcome following SCS. These findings are complex to extrapolate. Back pain hypoesthesia is one main criterion of the DN4 (Douleur Neuropathique en 4 questions) questionnaire [59], confirming that SCS is an appropriate tool to address neuropathic pain [10]. The positional exacerbation of pain might indicate that there is a mix of neuropathic and mechanical pain components, which are associated in this PSPS patient population, as opposed to continuous pain, preventing patients from participation in daily activities. This could be considered as an indirect marker of patient willingness to maintain a certain level of activity, and as detailed in our paper focusing on socio-professional status and patient activity [58], we found that SCS outcomes improved when patients were able to develop interesting coping strategies with pain, to pursue kinesiophobia and fight for functional capacity preservation. These results finally suggest that psychological evaluation and pain typology are important to characterize in patient selection prior to SCS implantation.

4.4. Study strengths and limitations

Apart from the originality of combining several machine learning models to reach our objectives, this study has several strengths, such as:
-   a double dataset, extracted from prospective comparative studies, comprising a large sample size, to reduce biases typically associated with this type of research and to ensure a maximal potential of generalization of our predictive models.
-   the multicenter nature of our sample also helps to ensure generalizability and applicability to clinical practice.
-   the development of a composite outcome, based on objective methods, which might lead to optimize patient satisfaction evaluation and therefore more precisely capture patient needs and expectations so as to define what should be considered as a positive outcome.

We also observed substantial limitations, which could be addressed by further research:
-   ***"Paresthesia intolerance"***

One of the reasons patients might fail a conventional tonic SCS trial is their inability to tolerate SCS-induced paresthesia. Patient potential intolerance to SCS-induced paresthesia was not taken into consideration in these study outcomes. In contrast, it has been shown that there is a high correlation between intolerance to tonic SCS and intolerance to TENS-induced paresthesia [44]. Therefore, combining a TENS trial with our machine learning model could serve as a noninvasive multiplexed screening tool prior to SCS permanent implantation for conventional SCS. However, when applying new paresthesia-free neurostimulation waveforms, such as high frequency, high density and BURST stimulation, this notion would no longer remain relevant to analyze.

-   ***"Country-dependent variety of SCS practices"***

Even though our data were extracted from several studies, including one multicenter national study with 12 centers, a main limitation would concern the relative homogeneity of practices across countries, such as lead choice, permanent trialing available in France and Europe, lead positioning, lead programming, depending on local cultures, which might not offer global extrapolations. An interesting way to overcome this limitation would be to pool and analyze data from large international registries, applying our machine learning model strategy, in further collaborative works.

- *"Large-scale validation"*

Although we have shown that our model's predictive power is superior to the screening-trial in this sample of 103 SCS implanted PSPS patients, it will remain necessary to conduct a more robust comparative study to assess the overall superiority of utilizing machine learning models as opposed to screening-trial on a larger cohort of patients. The evaluation criteria of such a study should include not only the comparative evaluation of lead screening trial vs mathematical models, but also medico-economical evaluations, extrapolated from models through healthcare-based medicine, and prospectively reported SCS implantation-related adverse events.

5. Conclusion

Machine learning and statistical models appear to show potential interest to put traditional SCS lead trial clinical outcomes into perspective with AI-based prediction of SCS efficacy. The regularized logistic regression, random forest and Gradient boosted trees models have demonstrated the best performance and provided a good model fit to the testing data and a relatively good performance on training data. Almost all the proposed models show a better prediction power than lead trial outcomes for this specific population of PSPS patients. On a general aspect, these results reinforce the need for AI-based predictive medicine, with no other ambition than helping physicians to optimize their clinical choices, based on a synergistic mathematical approach. SCS applications are at an early stage and by essence, remain limited. This deserves further study, based on high-fidelity composite multi-dimensional pain assessment and data extracted from large cohorts of implanted patients, to take this debate to another level. More specifically, in echo with recent clinical data, this study can be depicted as a seminal mathematical substrate to reconsider SCS lead trial utility.

References

1. North, R.B.; Campbell, J.N.; James, C.S.; Conover-Walker, M.K.; Wang, H.; Piantadosi, S.; Rybock, J.D.; Long, D.M. Failed Back Surgery Syndrome: 5-Year Follow-up in 102 Patients Undergoing Repeated Operation. *Neurosurgery* **1991**, *28*, 685–690; discussion 690-691.

2. Schug, S.A.; Lavand'homme, P.; Barke, A.; Korwisi, B.; Rief, W.; Treede, R.-D.; IASP Taskforce for the Classification of Chronic Pain The IASP Classification of Chronic Pain for ICD-11: Chronic Postsurgical or Posttraumatic Pain. *Pain* **2019**, *160*, 45–52, doi:10.1097/j.pain.0000000000001413.

3. Christelis, N.; Simpson, B.; Russo, M.; Stanton-Hicks, M.; Barolat, G.; Thomson, S.; Schug, S.; Baron, R.; Buchser, E.; Carr, D.B.; et al. Persistent Spinal Pain Syndrome: A Proposal for Failed Back Surgery Syndrome and ICD-11. *Pain Med* **2021**, doi:10.1093/pm/pnab015.

4. Ganty, P.; Sharma, M. Failed Back Surgery Syndrome: A Suggested Algorithm of Care. *Br J Pain* **2012**, *6*, 153–161, doi:10.1177/2049463712470222.

5. Law, J.D.; Lehman, R.A.; Kirsch, W.M. Reoperation after Lumbar Intervertebral Disc Surgery. *J Neurosurg* **1978**, *48*, 259–263, doi:10.3171/jns.1978.48.2.0259.

6. Lehmann, T.R.; LaRocca, H.S. Repeat Lumbar Surgery. A Review of Patients with Failure from Previous Lumbar Surgery Treated by Spinal Canal Exploration and Lumbar Spinal Fusion. *Spine (Phila Pa 1976)* **1981**, *6*, 615–619.

7. Naiditch, N.; Billot, M.; Moens, M.; Goudman, L.; Cornet, P.; Le Breton, D.; Roulaud, M.; Ounajim, A.; Page, P.; Lorgeoux, B.; et al. Persistent Spinal Pain Syndrome Type 2 (PSPS-T2), a Social Pain? Advocacy for a Social Gradient of Health Approach to Chronic Pain. *Journal of Clinical Medicine* **2021**, *10*, 2817, doi:10.3390/jcm10132817.

8. Inoue, S.; Kamiya, M.; Nishihara, M.; Arai, Y.-C.P.; Ikemoto, T.; Ushida, T. Prevalence, Characteristics, and Burden of Failed Back Surgery Syndrome: The Influence of Various Residual Symptoms on Patient Satisfaction and Quality of Life as Assessed by a Nationwide Internet Survey in Japan. *J Pain Res* **2017**, *10*, 811–823, doi:10.2147/JPR.S129295.

9. Farber, S.H.; Han, J.L.; Elsamadicy, A.A.; Hussaini, Q.; Yang, S.; Pagadala, P.; Parente, B.; Xie, J.; Lad, S.P. Long-Term Cost Utility of Spinal Cord Stimulation in Patients with Failed Back Surgery Syndrome. *Pain Physician* **2017**, *20*, E797–E805.

10. Kumar, K.; Taylor, R.S.; Jacques, L.; Eldabe, S.; Meglio, M.; Molet, J.; Thomson, S.; O'Callaghan, J.; Eisenberg, E.; Milbouw, G.; et al. The Effects of Spinal Cord Stimulation in Neuropathic Pain Are Sustained: A 24-Month Follow-up of the Prospective Randomized Controlled Multicenter Trial of the Effectiveness of Spinal Cord Stimulation. *Neurosurgery* **2008**, *63*, 762–770; discussion 770, doi:10.1227/01.NEU.0000325731.46702.D9.

11. Taylor, R.S.; Ryan, J.; O'Donnell, R.; Eldabe, S.; Kumar, K.; North, R.B. The Cost-Effectiveness of Spinal Cord Stimulation in the Treatment of Failed Back Surgery Syndrome. *Clin J Pain* **2010**, *26*, 463–469, doi:10.1097/AJP.0b013e3181daccec.

12. Eldabe, S.S.; Taylor, R.S.; Goossens, S.; Bouche, B.; Gültuna, I.; Green, C.; Tinsley, J.; Luyet, P.; Buchser, E. A Randomized Controlled Trial of Subcutaneous Nerve Stimulation for Back Pain Due to Failed Back Surgery Syndrome: The SubQStim Study. *Neuromodulation* **2019**, *22*, 519–528, doi:10.1111/ner.12784.

13.   North, R.B.; Kidd, D.H.; Farrokhi, F.; Piantadosi, S.A. Spinal Cord Stimulation versus Repeated Lumbosacral Spine Surgery for Chronic Pain: A Randomized, Controlled Trial. *Neurosurgery* **2005**, *56*, 98–107, doi:10.1227/01.NEU.0000144839.65524.E0.

14.   Rigoard, P.; Basu, S.; Desai, M.; Taylor, R.; Annemans, L.; Tan, Y.; Johnson, M.J.; Van den Abeele, C.; North, R.; PROMISE Study Group Multicolumn Spinal Cord Stimulation for Predominant Back Pain in Failed Back Surgery Syndrome Patients: A Multicenter Randomized Controlled Trial. *Pain* **2019**, *160*, 1410–1420, doi:10.1097/j.pain.0000000000001510.

15.   Ounajim, A.; Billot, M.; Louis, P.-Y.; Slaoui, Y.; Frasca, D.; Goudman, L.; Roulaud, M.; Naiditch, N.; Lorgeoux, B.; Baron, S.; et al. Finite Mixture Models Based on Pain Intensity, Functional Disability and Psychological Distress Composite Assessment Allow to Identify Two Distinct Classes of Persistent Spinal Pain Syndrome after Surgery Patients Related to Their Quality of Life. **2021**, doi:10.20944/preprints202108.0527.v1.

16.   National Institute for Health and Clinical Excellence Spinal Cord Stimulation for Chronic Pain of Neuropathic or Ischaemic Origin. *Retrieved from https://www.nice.org.uk/guidance/ta159*.

17.   Centers for Medicare & Medicaid Services National Coverage Determination (NCD) for Electrical Nerve Stimulators. 160:7.

18.   Rigoard, P.; Billot, M.; Ingrand, P.; Durand-Zaleski, I.; Roulaud, M.; Peruzzi, P.; Dam Hieu, P.; Voirin, J.; Raoul, S.; Page, P.; et al. How Should We Use Multicolumn Spinal Cord Stimulation to Optimize Back Pain Spatial Neural Targeting? A Prospective, Multicenter, Randomized, Double-Blind, Controlled Trial (ESTIMET Study). *Neuromodulation* **2020**, doi:10.1111/ner.13251.

19.   HAUTE AUTORITE DE SANTE Evaluation Des Systèmes Implantables de Neurostimulation Médullaire. *https://www.hassante.fr/upload/docs/application/pdf/2014-04/fiche_de_bon_usage_neurostimulation_2014-04-03_10-38-24_855.pdf*.

20.   Deer, T.R.; Lamer, T.J.; Pope, J.E.; Falowski, S.M.; Provenzano, D.A.; Slavin, K.; Golovac, S.; Arle, J.; Rosenow, J.M.; Williams, K.; et al. The Neurostimulation Appropriateness Consensus Committee (NACC) Safety Guidelines for the Reduction of Severe Neurological Injury. *Neuromodulation: Technology at the Neural Interface* **2017**, *20*, 15–30, doi:https://doi.org/10.1111/ner.12564.

21.   Thomson, S.; Huygen, F.; Prangnell, S.; De Andrés, J.; Baranidharan, G.; Belaïd, H.; Berry, N.; Billet, B.; Cooil, J.; De Carolis, G.; et al. Appropriate Referral and Selection of Patients with Chronic Pain for Spinal Cord Stimulation: European Consensus Recommendations and E-health Tool. *Eur J Pain* **2020**, *24*, 1169–1181, doi:10.1002/ejp.1562.

22.   Van Buyten, J.P.; Van Zundert, J.; Vueghs, P.; Vanduffel, L. Efficacy of Spinal Cord Stimulation: 10 Years of Experience in a Pain Centre in Belgium. *Eur J Pain* **2001**, *5*, 299–307, doi:10.1053/eujp.2001.0249.

23.   Rigoard, P.; Ounajim, A.; Goudman, L.; Louis, P.-Y.; Yousri, S.; Roulaud, M.; Bouche, B.; Wood, C.; Page, P.; Lorgeoux, B.; et al. A Novel Multi-Dimensional Clinical Response Index Dedicated to Improve Pain Global Assessment in Patients with Persistent Spinal Pain Syndrome after Spinal Surgery, Based on a Real-Life Prospective Multicentric Study (PREDIBACK) and Machine Learning Techniques. **2021**, doi: 10.20944/preprints202108.0545.v1.

24.   Eldabe, S.; Duarte, R.V.; Gulve, A.; Thomson, S.; Baranidharan, G.; Houten, R.; Jowett, S.; Sandhu, H.; Chadwick, R.; Brookes, M.; et al. Does a Screening Trial for Spinal Cord Stimulation in Patients with Chronic Pain of Neuropathic Origin Have Clinical Utility and Cost-Effectiveness (TRIAL-STIM)? A Randomised Controlled Trial. *Pain* **2020**, *161*, 2820–2829, doi:10.1097/j.pain.0000000000001977.

25.   Weinand, M.E.; Madhusudan, H.; Davis, B.; Melgar, M. Acute vs. Prolonged Screening for Spinal Cord Stimulation in Chronic Pain. *Neuromodulation* **2003**, *6*, 15–19, doi:10.1046/j.1525-1403.2003.03002.x.

26. Simopoulos, T.; Sharma, S.; Aner, M.; Gill, J.S. A Temporary vs. Permanent Anchored Percutaneous Lead Trial of Spinal Cord Stimulation: A Comparison of Patient Outcomes and Adverse Events. *Neuromodulation* **2018**, *21*, 508–512, doi:10.1111/ner.12687.

27. Bendersky, D.; Yampolsky, C. Is Spinal Cord Stimulation Safe? A Review of Its Complications. *World Neurosurg* **2014**, *82*, 1359–1368, doi:10.1016/j.wneu.2013.06.012.

28. North, R.; Desai, M.J.; Vangeneugden, J.; Raftopoulos, C.; Van Havenbergh, T.; Deruytter, M.; Remacle, J.; Shipley, J.; Tan, Y.; Johnson, M.J.; et al. Postoperative Infections Associated With Prolonged Spinal Cord Stimulation Trial Duration (PROMISE RCT). *Neuromodulation* **2020**, *23*, 620–625, doi:10.1111/ner.13141.

29. Strauss, I.; Taha, K.; Krishna, V.; Hodaie, M. Younger Age Predicts Greater Effectiveness of Spinal Cord Stimulation for Chronic Pain. *Acta Neurochir (Wien)* **2016**, *158*, 999–1003, doi:10.1007/s00701-016-2753-0.

30. Celestin, J.; Edwards, R.R.; Jamison, R.N. Pretreatment Psychosocial Variables as Predictors of Outcomes Following Lumbar Surgery and Spinal Cord Stimulation: A Systematic Review and Literature Synthesis. *Pain Med* **2009**, *10*, 639–653, doi:10.1111/j.1526-4637.2009.00632.x.

31. Caylor, J.; Reddy, R.; Yin, S.; Cui, C.; Huang, M.; Huang, C.; Rao, R.; Baker, D.G.; Simmons, A.; Souza, D.; et al. Spinal Cord Stimulation in Chronic Pain: Evidence and Theory for Mechanisms of Action. *Bioelectronic Medicine* **2019**, *5*, 12, doi:10.1186/s42234-019-0023-1.

32. Mekhail, N.; Mehanny, D.S.; Armanyous, S.; Costandi, S.; Saweris, Y.; Azer, G.; Bolash, R. Choice of Spinal Cord Stimulation versus Targeted Drug Delivery in the Management of Chronic Pain: A Predictive Formula for Outcomes. *Reg Anesth Pain Med* **2020**, *45*, 131–136, doi:10.1136/rapm-2019-100859.

33. Campbell, C.M.; Jamison, R.N.; Edwards, R.R. Psychological Screening/Phenotyping as Predictors for Spinal Cord Stimulation. *Curr Pain Headache Rep* **2013**, *17*, 307, doi:10.1007/s11916-012-0307-6.

34. Bendinger, T.; Plunkett, N.; Poole, D.; Turnbull, D. Psychological Factors as Outcome Predictors for Spinal Cord Stimulation. *Neuromodulation* **2015**, *18*, 465–471; discussion 471, doi:10.1111/ner.12321.

35. Sparkes, E.; Duarte, R.V.; Mann, S.; Lawrence, T.R.; Raphael, J.H. Analysis of Psychological Characteristics Impacting Spinal Cord Stimulation Treatment Outcomes: A Prospective Assessment. *Pain Physician* **2015**, *18*, E369-377.

36. Taylor, R.S.; Desai, M.J.; Rigoard, P.; Taylor, R.J. Predictors of Pain Relief Following Spinal Cord Stimulation in Chronic Back and Leg Pain and Failed Back Surgery Syndrome: A Systematic Review and Meta-Regression Analysis. *Pain Pract* **2014**, *14*, 489–505, doi:10.1111/papr.12095.

37. Kumar, K.; Rizvi, S.; Nguyen, R.; Abbas, M.; Bishop, S.; Murthy, V. Impact of Wait Times on Spinal Cord Stimulation Therapy Outcomes. *Pain Pract* **2014**, *14*, 709–720, doi:10.1111/papr.12126.

38. Goudman, L.; Van Buyten, J.-P.; De Smedt, A.; Smet, I.; Devos, M.; Jerjir, A.; Moens, M. Predicting the Response of High Frequency Spinal Cord Stimulation in Patients with Failed Back Surgery Syndrome: A Retrospective Study with Machine Learning Techniques. *J Clin Med* **2020**, *9*, 4131, doi:10.3390/jcm9124131.

39. Brooks, R. EuroQol: The Current State of Play. *Health Policy* **1996**, *37*, 53–72, doi:10.1016/0168-8510(96)00822-6.

40. Fairbank, J.C.; Couper, J.; Davies, J.B.; O'Brien, J.P. The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy* **1980**, *66*, 271–273.

41. Montgomery, S.A.; Asberg, M. A New Depression Scale Designed to Be Sensitive to Change. *Br J Psychiatry* **1979**, *134*, 382–389, doi:10.1192/bjp.134.4.382.

42. Marola, O.; Cherala, R.; Prusik, J.; Kumar, V.; Fama, C.; Wilock, M.; Crimmins, J.; Pilitsis, J.G. BMI as a Predictor of Spinal Cord Stimulation Success in Chronic Pain Patients. *Neuromodulation* **2017**, *20*, 269–273, doi:10.1111/ner.12482.

43. van Eijs, F.; Smits, H.; Geurts, J.W.; Kessels, A.G.H.; Kemler, M.A.; van Kleef, M.; Joosten, E.A.J.; Faber, C.G. Brush-Evoked Allodynia Predicts Outcome of Spinal Cord Stimulation in Complex Regional Pain Syndrome Type 1. *Eur J Pain* **2010**, *14*, 164–169, doi:10.1016/j.ejpain.2009.10.009.

44. Mathew, L.; Winfree, C.; Miller-Saultz, D.; Sonty, N. Transcutaneous Electrical Nerve Stimulator Trial May Be Used as a Screening Tool Prior to Spinal Cord Stimulator Implantation. *Pain* **2010**, *150*, 327–331, doi:10.1016/j.pain.2010.05.018.

45. Goudman, L.; De Smedt, A.; Eldabe, S.; Rigoard, P.; Linderoth, B.; De Jaeger, M.; Moens, M.; Consortium, D. High-Dose Spinal Cord Stimulation for Patients with Failed Back Surgery Syndrome: A Multicenter Effectiveness and Prediction Study. *PAIN* **2020**, *Articles in Press*, doi:10.1097/j.pain.0000000000002035.

46. Friedman, J.H.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **2010**, *33*, 1–22, doi:10.18637/jss.v033.i01.

47. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support Vector Machines. *IEEE Intelligent Systems and their Applications* **1998**, *13*, 18–28, doi:10.1109/5254.708428.

48. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F.; C++-code), C.-C.C. (libsvm; C++-code), C.-C.L. (libsvm *E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*; 2020;

49. Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian Network Classifiers. *Machine Learning* **1997**, *29*, 131–163, doi:10.1023/A:1007465528199.

50. Haykin, S. *Neural Networks: A Comprehensive Foundation*; 2nd ed.; Prentice Hall PTR: USA, 1998; ISBN 978-0-13-273350-2.

51. Arnold, T.B. KerasR: R Interface to the Keras Deep Learning Library. *J. Open Source Softw.* **2017**, doi:10.21105/joss.00296.

52. Therneau, T.; Atkinson, B.; port, B.R. (producer of the initial R.; maintainer 1999-2017) *Rpart: Recursive Partitioning and Regression Trees*; 2019;

53. Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32, doi:10.1023/A:1010933404324.

54. Wright, M.N.; Wager, S.; Probst, P. *Ranger: A Fast Implementation of Random Forests*; 2020;

55. Friedman, J.H. Stochastic Gradient Boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378, doi:10.1016/S0167-9473(01)00065-2.

56. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* **2016**, 785–794, doi:10.1145/2939672.2939785.

57. Rudiger, J.; Thomson, S. Infection Rate of Spinal Cord Stimulators After a Screening Trial Period. A 53-Month Third Party Follow-Up. *Neuromodulation: Technology at the Neural Interface* **2011**, *14*, 136–141, doi:https://doi.org/10.1111/j.1525-1403.2010.00317.x.

58. Naiditch, N.; Billot, M.; Goudman, L.; Cornet, P.; Roulaud, M.; Ounajim, A.; Page, P.; Lorgeoux, B.; Baron, S.; Nivole, K.; et al. Professional Status of Persistant Spinal Pain Syndrome Patients after Spinal Surgery (PSPS-T2): What Does Really Matter? A pro-Spective Study Introducing the Concept of "Adapted Profession-al Activity" Infering from Clinical, Psychological and Social in-Fluence. *Journal of Clinical Medicine* 2021.

59. Bouhassira, D.; Attal, N.; Fermanian, J.; Alchaar, H.; Gautron, M.; Masquelier, E.; Rostaing, S.; Lanteri-Minet, M.; Collin, E.; Grisart, J.; et al. Development and Validation of the Neuropathic Pain Symptom Inventory. *PAIN* **2004**, *108*, 248–257, doi:10.1016/j.pain.2003.12.024.