UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics

Master's thesis

The effectiveness of High Dose Spinal Cord Stimulation on disability: a longitudinal analysis

Lisa Goudman

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Epidemiology & Public Health Methodology

SUPERVISOR :
Prof. dr. Geert MOLENBERGHS

SUPERVISOR :
Prof. dr. Maarten MOENS

2019
2020

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics

Master's thesis

The effectiveness of High Dose Spinal Cord Stimulation on disability: a longitudinal analysis

Lisa Goudman

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Epidemiology & Public Health Methodology

SUPERVISOR :

Prof. dr. Geert MOLENBERGHS

SUPERVISOR :

Prof. dr. Maarten MOENS

Acknowledgements

Presumably all data analysists, statisticians, epidemiologists and all other persons in the field of *methodology* are familiar with the word 'sensitivity analysis'. However, how to perform a sensitivity analysis and what software to use for this type of analysis might be less known. Personally, I only had a rough idea about these analyses, wherefore I was not familiar with sensitivity analysis at all at the beginning of the second semester. This thesis really challenged me to keep discovering new types of analyses, for which I'm really grateful to prof. Molenberghs for guiding me into the right direction. I would like to thank prof. Molenberghs for his patience, guidance, help, quick answers and feedback on this work.

This thesis is performed in collaboration with prof. Moens who conducted a multicenter registry to evaluate the effectiveness of Spinal Cord Stimulation in chronic pain patients. I would like to thank prof. Moens for recruiting these patients and for providing us full access to the data. Additionally, I like to thank prof. Moens for his supportive words when I got struggled with this thesis.

Finally, I would like to thank everyone who encouraged me to persevere with this program.

Contents

Abstract

Introduction. Since its first use in humans in 1967, Spinal Cord Stimulation (SCS) has been established as an effective therapy to treat chronic pain conditions. More recently, new waveforms and frequencies have changed the paradigm of standard SCS to address SCS long-term failures and to optimize therapy outcomes. The use of High Dose SCS (HD-SCS) has drastically increased during the last years, with positive preliminary results in terms of pain relief. However, to increase quality of life of chronic pain patients, one should not only strive towards a pain reduction but also achieve a decrease in disability or a reduction in opioid consumption. Therefore, the primary aim of this study was to evaluate the effectiveness of HD-SCS on disability in patients with Failed Back Surgery Syndrome (FBSS).

Methods. One hundred eighty-five patients with FBSS were included in this study. Disability and pain intensity scores were evaluated at baseline (before receiving SCS) and after 1, 3 and 12 months of neurostimulation with HD-SCS. During the second, third and fourth visit respectively data of 130, 114 and 90 patients was available. Longitudinal mixed models were used to evaluate disability over time. Afterwards a tipping point sensitivity analysis was performed.

Results. HD-SCS significantly decreased disability scores in patients with FBSS. The sensitivity analysis revealed that the shift parameter was 17. Thus, the conclusion concerning the time effect under the 'Missing at random' mechanism is reserved when the shift parameter for the disability score is 17.

Discussion. Patients with FBSS benefit from HD-SCS not only in terms of pain relief but also to decrease disability. From a clinical point of view, a shift of 17 points on disability is not very plausible, wherefore we are tended to accept the conclusions drawn under 'Missing at random'.

1 Introduction

Since its first use in humans in 1967 [35], Spinal Cord Stimulation (SCS) has been established as an effective therapy to treat a wide variety of chronic pain conditions. One of the conditions in which SCS is often applied as treatment is Failed Back Surgery Syndrome (FBSS). This condition is characterized by persistent back and/or leg pain of unknown origin either despite surgical intervention or appearing after surgical intervention for spinal pain [4]. The incidence of patients that will develop FBSS after lumbar spinal surgery is estimated in the range of 10-40%, depending on the exact type of surgery [5, 34].

SCS involves the implantation of an epidural electrode, which is connected through extensions with a subcutaneous implanted pulse generator [35]. Electrical pulses at different frequencies are generated and delivered to the spinal cord to elicit paresthesia in the painful area [11]. The goal of SCS is not to cure patients but rather to make chronic pain tolerable, with benefits on functionality and health-related quality of life [38, 30].

Initially, standard SCS was provided whereby patients are experiencing paresthesia in the painful areas. Over the last decade, several new waveforms and frequencies were introduced which are not inducing paresthesia anymore [18]. One of those new paresthesia-free stimulation paradigms, launched in 2016, is High Dose SCS (HD-SCS). HD-SCS entails an increase in frequency and pulse width, along with a reduced amplitude, when compared to conventional SCS [21]. Despite the absence of an exact definition for the stimulation parameters of HD-SCS, the delivery of energy to neural tissue is the key concept behind this paradigm [41]. The percentage of active stimulation during a pulse cycle can be increased up to 20–25% for the maximal available settings, at a subsensory mode [21, 18]. The first reports on HD-SCS were promising with benefits in terms of pain relief [26, 6, 12].

When evaluating the success of a treatment in the field of neuromodulation, the most prominent outcome measurement is a reduction in pain intensity. However, it has previously been demonstrated that achieving a pre-treatment goal of 'reducing pain' contributes very little to patient satisfaction in chronic disabled back and/or neck pain patients [13]. Moreover, achieving "functional goals" was more important for patient satisfaction than a reduction in self-reported pain [13]. Additionally, a qualitative exploration towards patients' expectations on SCS indicated that patients have more expectations than only obtaining pain relief [14]. These studies, combined with the recent call in the SCS literature to focus on a combination of several outcome measurements [25], clearly demonstrates that we should redefine the definition of a successful treatment in SCS.

Disability is one of the factors that can be proposed as additional self-reporting measurement for evaluating the treatment effect of SCS. One of the most frequently used questionnaires to evaluate disability within patients with chronic low back pain, is the Oswestry Disability Index (ODI) [8]. This questionnaire, initially develop by O'Brien in 1976, consists of ten sections measuring pain intensity, personal care, lifting, walking, sitting, standing, sleeping, sex life, social life, and travelling. A moderate correlation has been reported between the ODI the 36-Item Short Form Survey as a measure of health status [10]. Furthermore, changes in disability will easily affect other outcome domains, among which quality of life and work status [27].

To evaluate the effect of HD-SCS from a more holistic approach, disability was used as outcome measure in this longitudinal study, in which data was available up to 12 months after the initiation of the treatment. A specific feature of longitudinal data is that they are clustered, i.e. clusters are composed of the repeated measurements obtained from a single individual at different visits, whereby observations in a cluster typically exhibit a positive correlation [9]. Due to the repeated measures, mixed models were applied to perform a longitudinal analysis. This enabled the evaluation of the within-subject changes in the response over time. Afterwards, a sensitivity analysis was performed to estimate the robustness of the main results. The general aim of this study was to evaluate the long-term effectiveness of HD-SCS in patients with FBSS on disability.

2 Data

For this thesis, we used data from the "Discover" project (Clinicaltrials.gov NCT02787265). This prospective, multicenter registry was designed to assess the effectiveness of HD-SCS in patients with FBSS. Patients were recruited between October 2016 and August 2018 in 15 Belgian neuromodulation centers and 1 center in France, all with ample HD-SCS experience. All inlcuded patients underwent a baseline visit which was scheduled before SCS implantation. After a SCS trial period of 4 weeks, a definitive SCS was implanted (minimal invasive surgical intervention). All patients were implanted with a RestoreSensor, Intellis or PrimeAdanced IPG (Minneapolis, MN, USA) and received HD-SCS with a pulse density of 25% (500 Hz and 500 sec of pulse) in case of the RestoreSensor or Intellis and 11.7% (450 Hz and 130 sec of pulse width) in case of a PrimeAdvanced IPG. Subsequently, three visits took place after respectively one month, three months and twelve months of HD-SCS.

At the first visit, patient demographics were recorded. At all study visits, disbility, pain intensity for leg pain and pain intensity for back pain were recorded. Age was used as a categorical variable with three age categories namely young patients (25-45 years), middle aged patients (46-65 years) and older patients (66-85 years). For disability, the Oswestry Disability Index (ODI) was used. The total score on this questionnaire is ranging from 0

to 100, with higher values representing more disability. Pain intensity was measured with the Numeric Rating Scale (NRS) for both leg and back pain separately. The NRS ranges from 0 to 10 whereby 0 represents no pain and 10 the worst imaginable pain.

According to the protocol of this study, disability is recorded at 4 visits for each patients. Due to the repeated data, longitudinal mixed models were used to evaluate the effectiveness of HD-SCS on disability in patients with FBSS.

3 Longitudinal mixed models

This part is based on the books of Fitzmaurice [9] and Molenberghs & Verbeke [22, 40].

3.1 The model

Let Y_{ij} denote the response variable for the i^{th} subject (i= 1,...,N) at the j^{th} time point (j= 1,...,n_i). Let $\mathbf{Y_i}$ be an n_i-dimensional vector of all repeated measurements for subject i (i.e. $\mathbf{Y_i} = (Y_{i1}, Y_{i2}, ..., Y_{in_i})'$).

The general form of a linear mixed model can be written as follows:

$$\begin{cases} \mathbf{Y_i} = X_i\beta + Z_i\mathbf{b_i} + \epsilon_i \\ \mathbf{b_i} \sim N(0, D) \\ \epsilon_\mathbf{i} \sim N(0, \Sigma_i) \\ \mathbf{b_1}, ..., \mathbf{b_N}, \epsilon_\mathbf{1}, ..., \epsilon_\mathbf{N} \text{ independent} \end{cases}$$

whereby β is a (p x 1) vector of fixed effects, $\mathbf{b_i}$ is a (q x 1) vector of random effects, X_i is a (n_i x p) matrix of covariates, and Z_i is a (n_i x q) matrix of covariates (often called the design matrix). D is a (q x q) covariance matrix. Σ is a (n_i x n_i) covariance matrix. Note that Σ only depends on i through n_i. $\epsilon_\mathbf{i}$ is an n_i-dimensional vector of residual components. Σ is often equal to $\sigma^2 I_{n_i}$ where In_i denotes a (n_i x n_i) identity matrix.

Furthermore, a distinction should be made between a conditional and a marginal model. The conditional (or subject-specific) mean of Y_i, given b_i, is

$$E(Y_i|b_i) = X_i\beta + Z_ib_i$$

The marginal (or population-averaged) mean of Y_i, averaged over the distribution of the random effects b_i, is:

$$\begin{aligned} E(Y_i) &= \mu_i \\ &= E\{E(Y_i|b_i)\} \\ &= E(X_i\beta + Z_ib_i) \\ &= X_i\beta + Z_iE(b_i) \\ &= X_i\beta \end{aligned}$$

The marginal density function of Y_i is given by:

$$f(\mathbf{y_i}) = \int f(\mathbf{y_i}|\mathbf{b_i}) f(\mathbf{b_i}) \mathrm{d}\mathbf{b_i}$$

whereby $f(\mathbf{y_i}|\mathbf{b_i})$ and $f(\mathbf{b_i})$ are density functions. As such, the marginal density function of $\mathbf{Y_i}$ has a mean vector $X_i\beta$ and covariance matrix $V_i = Z_i D Z_i' + \Sigma_i$. Thus, within a linear mixed effects model, β's (fixed effects) are assumed to be similar for all individuals and have an interpretation on the population-averaged level. The b_i vector contains subject-specific regression coefficients (i.e. random effects). When combining b_i with the fixed effects, they are describing the mean response profile for any individual.

Given the full distributional assumptions about the vector of responses Y_i, maximum likelihood (ML) is used for estimation. The main idea behind this approach is to select those values as estimates for β and Σ_i that are most likely for the data that are actually observed. Estimation of β and Σ_i proceeds by maximizing the likelihood function, i.e. the probability of the response variables evaluated at the fixed set of observed values and regarded as functions of β and Σ_i. The values that maximize the likelihood function are called maximum likelihood estimates of β and Σ_i, denoted as $\widehat{\beta}$ and $\widehat{\Sigma_i}$.

3.2 Statistical inference

Let L be a single row vector of known weights with the null hypothesis (H_0) : $\mathrm{L}\beta = 0$ and alternative hypothesis (H_A): $\mathrm{L}\beta \neq 0$. To test H_0 versus H_A the Wald statistic can be compared to a standard normal distribution:

$$Z = \frac{L\widehat{\beta}}{\sqrt{L\widehat{\mathrm{Cov}}(\widehat{\beta})L'}}$$

whereby $L\mathrm{Cov}(\widehat{\beta})L'$ is a single value and the square root is the estimate of the standard error for $L\widehat{\beta}$. When L has more rows (i.e. r rows which are representing r contrasts of interest), a multivariate Wald test is used to compare H_0 versus H_A which has a χ^2 distribution with r degrees of freedom:

$$W^2 = (L\widehat{\beta})'\{L\widehat{\mathrm{Cov}}(\widehat{\beta})L'\}^{-1}(L\widehat{\beta})$$

Alternatively, H_0 : $\mathrm{L}\beta = 0$ versus H_A: $\mathrm{L}\beta \neq 0$ could be evaluated with a Likelihood Ratio test (LRT). This test compares the maximized log-likelihoods for two models namely a full model (defined as an unconstrained model) and a reduced model (constrained model with $\mathrm{L}\beta = 0$). The reduced model is nested within the full model. The larger the difference between maximized log-likelihoods, the stronger the evidence that the reduced model is not adequate. Formally:

$$G^2 = 2(\hat{\mathrm{l}}_{\mathrm{full}} - \hat{\mathrm{l}}_{\mathrm{red}})$$

should be compared to a χ^2 distribution with degrees of freedom equal to the difference between the number of parameters in the full and reduced model.

Restricted (or residual) maximum likelihood (REML) estimation should be applied for estimating Σ_i. The main concept of REML estimation is to separate the data that is used for estimating Σ_i from β in order to eliminate β from the likelihood estimation of Σ_i. The REML estimator will be less seriously biased than the ML estimator for Σ_i. If the sample size is substantially larger than p (the dimension of β), then the difference between ML and REML estimation becomes less important. This approach should be recommended for comparing nested models for the covariance but not to compare nested regression models for the mean. When dealing with non-nested covariance models, Akaike Information Criterion (AIC) could be used. Among several non-nested competing models for the covariance, the model which minimizes the following expression should be selected.

$AIC =$ -2 (maximized REML log-likelihood) + 2 (number of covariance parameters)

In this framework, the objective is to select a model that has a good fit to the data and a model that is parsimonious. This is obtained by extracting a penalty for the estimation of each additional covariance parameter. Thus, AIC can be used to compare models with the same fixed effects but with different models for the covariance.

4 Methodology

To gain insight in the data, univariate analyses and different plots to evaluate correlations and (co)variances were constructed. Concerning the model building, we started by fitting a nearly saturated mean model that includes all main effects, two way- and three way interaction terms. We started the model building with an unstructured covariance matrix, allowing different variances on each visit and different correlations between all combinations of visits. The necessity of random slopes and/or random intercepts was evaluated by Restricted Maximum Likelihood (REML) estimation. Splines were also considered (REML estimation). Secondly, the unstructured covariance matrix was compared with other covariance structures. Model selection was performed with AIC values. If the unstructured covariance matrix model did not differed significantly from a model with more assumptions, we replaced it. Thirdly, the fixed part of the model was simplified by dropping unnecessary predictors using LR tests, starting from the interaction terms. Deletion of a predictor was allowed if it does not affect the model ($p>0.05$). If a higher order interaction term needed to be included, the lower interaction terms and main effects remained in the model as well. All statistical analyses were performed in SAS 9.4 with PROC MIXED.

5 Results

5.1 Primary analysis

5.1.1 Descriptive statistics

In this study, 89 males (48.1%) and 96 females (51.9%) were included with a mean age of 54 (SD 12.01) years. The mean ODI score at baseline was 56.99 (SD 14.97), 31.26 (SD 17.58) at 1 month, 30.64 (SD 18.52) at 3 months and 33.34 (SD 16.86) at 12 months. At the first visit, data of 185 patients was available for the outcome variable. During the second, third and fourth visit respectively data of 130, 114 and 90 patients was available.

5.1.2 Exploratory data analysis

Overall, there seemed to be a decrease in average ODI score over time (Figure 1). There seems to be a lower variability in ODI score at baseline compared to the follow-up visits. The ODI score seems to decrease very fast from baseline to one month of SCS, and afterwards a more stable ODI score seems to be present. Presumably, a model with a knot at the first months would be an option to investigate.

When plotting the ODI score at baseline, there seems to be a higher ODI score for females compared to males. A decrease in ODI score is visible in both groups between baseline and the first visit. Over time, we do observe a difference between males and females whereby females seem to demonstrate a linear time effect from 1 month to 12 months and males a slight increase in ODI score from 3 months up to 12 months (Appendix A1). When plotting the different age categories in function of the ODI, the ODI score in the group with a middle age category seems to be the lowest at all time points (Appendix A2).

The individual profiles are plotted in Figure 2, which clearly confirm the fast decrease in ODI score during the first month. Based on these profiles, a model with a random intercept seems very plausible.

The scatterplot matrix (Appendix A3) and variance/covariance parameter estimates (Table 1) suggest a decaying correlation with increasing visit lags. Therefore, both a compound symmetry and autoregressive covariance matrix (constant variance) as well as a heterogeneous Toeplitz, heterogeneous autoregressive and heterogeneous compound symmetry covariance matrix (difference in variance) were fitted.

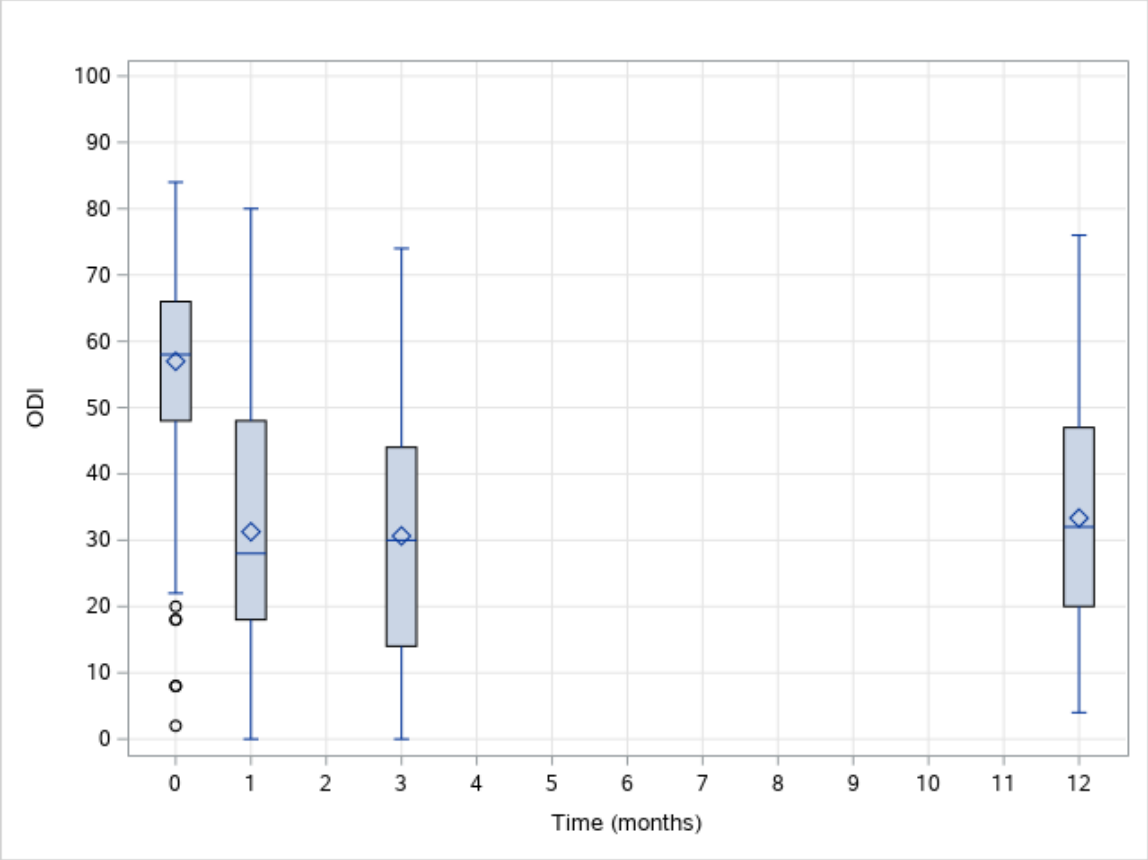


Figure 1: *Distribution of ODI score per visit. The mean is represented by the square, the horizontal line is representing the median ODI score at each visit.*

5.1.3 Model building

We started with a nearly saturated model for the mean with a unstructured covariance matrix. The data revealed a fast decrease in ODI score up to 1 month, where after a slow increase from the first month onward was present (Figure 1). Therefore, a knot was assumed at the first visit thanks to the creation of an additional variable 'Time1' which took the value zero when the observation occurred before the first month. Otherwise, the new variable took the value of the current month minus one. The model with random varying slopes was performing better than a model without the additional slope, wherefore we decided to keep the extra knot in the model. This was tested with a LR test statistic with a mixture of Chi-squared distributions with 2 and 3 degrees of freedom (LR=10.63, p=0.009). Next, we compared a model with a single covariance structure to a model where there is a different structure per sex and age category using a LR test. It was not necesarry to use a model with different covariance structures per sex compared to a model without

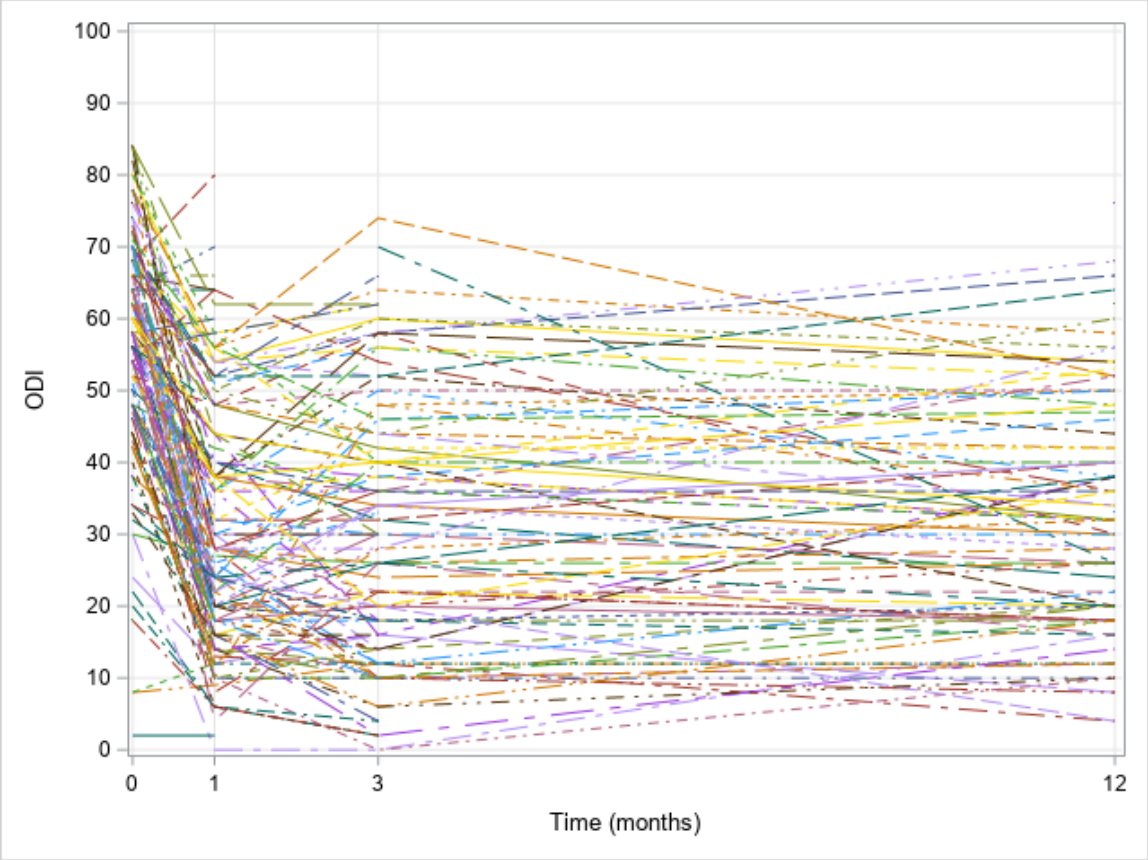


Figure 2: *Individual profile plots.*

different structures (LR=15.15, df=10, p=0.13). Similarly for age categories, a model with different covariance structures and a model without different structures were fitted, indicating that the model with different structures is favored (LR=39.14, df=20, p=0.006). Additionally, we controlled whether a random slope and random intercept model would be defensible compared to a random intercept model using REML estimation. The -2 Res Log Likelihood equaled 3751.34 and 3775.52 respectively for a model with random intercept and random slopes versus a random intercept model only, resulting in a Likelihood Ratio test statistic of 24.19 which was compared to a Chi-squared distribution with a mixture of 9 and 4 degrees of freedom (p=0.002). Therefore, a model with random slopes will be used.

Due to the covariance that decreases with larger differences between visits, a heterogeneous Toeplitz (TOEPH), heterogeneous compound symmetry (CSH), heterogeneous autoregressive (ARH(1)), autoregressive (AR(1)) and compound symmetry (CS) covariance matrix were fitted. The AR(1) model had the lowest AIC (3780.5), followed

by the UN (3783.3), CS (3784.7), ARH(1) (3790.0), TOEPH (3792.3) and CSH (3796.4). Based on the AIC, we decided to replace the unstructured covariance matrix by a model with less parameters namely AR(1).

Month	0	1	3	12
0	172.93			
1	100.94	169.07		
3	90.95	127.97	189.51	
12	99.35	111.02	138.35	162.34

Table 1: Estimated covariance parameters of the nearly saturated model with an unstructured covariance matrix. Diagonal elements represent the variances, while off-diagonal elements depict covariances.

Continuing with the AR(1) covariance matrix, the model was reduced by excluding the least significant predictors. None of the three-way interaction terms was necessary to be included in the model, wherefore they were all excluded (LRT=32.03, df=25 p=0.16). Starting from a model with all two-way interaction terms, it was possible to remove all two-way interaction terms (LRT=55.67, df=44 p=0.11). Subsequently, we sequentially removed the least significant term from the model until no further simplifications were possible anymore. This has lead to a final model with a random intercept and two random varying slopes, four main effects namely back pain intensity, leg pain intensity, time and time1. Additionally, different covariance matrices were allowed per age category. The regression coefficient estimates of the final model are presented in Table 2. QQ plots for the raw and scaled residuals were constructed (Appendix 4). There were no real departures from normality visible and no abundant outlying observations could be detected. The residuals had a symmetrical distribution around zero and no sign of heteroscedasticity were visible.

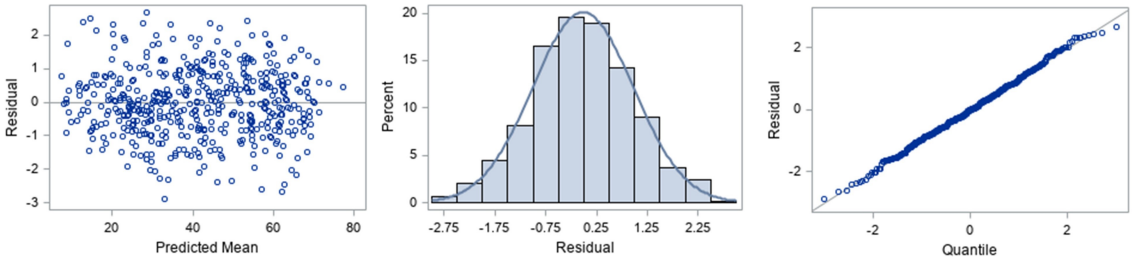


Figure 3: *Scatterplot, histogram and QQ plot for scaled residuals.*

5.1.4 Final model

At the first visit, the average ODI score is 25.05 ([20.43;29.66], p< 0.001) for a patient with a pain intensity score of 0 for both back and leg pain. For 95% of the patients, the average ODI score before treatment varies between 4.83 and 45.27 ($25.05 \pm 1.96 * \sqrt{(106.39)}$). Per unit increase in NRS back pain score, the average ODI score will increase with 2.32 ([1.82;2.81], F = 86.73, p< 0.001). For each unit increase in NRS leg pain score, the average ODI score will increase with 1.87 units ([1.44;2.30], F = 75.38, p< 0.001). There is a monthly decrease of 7.68 ([4.98;10.39], F = 31.58, p< 0.001) in average ODI score during the first month. For 95% of the patients, the average visit change in ODI score before the first visit varies up to 14.68 units away from the population mean. The percentage of patients that is experiencing an average decrease in ODI score before the first follow-up visit 84.7%. After the first visit, an increase of 7.61 ([4.84;10.38], F = 29.66, p< 0.001) in ODI score is revealed per visit. For 95% of the patients, the average change in evolution per visit after versus before the first month varies up to 14.98 units away from the population mean. The AR(1) correlation parameter of 0.6639 indicates that, within a middle aged subject, the correlation between two visits that are 1 time unit apart is 0.6639. The correlation within older subjects and young subjects is respectively 0.18 and 0.35 between two visits that are 1 time unit apart.

Variable	Regression estimates	SE	95% CI	Type III test
Intercept	25.05	2.34	[20.43 to 29.66]	p<0.001
NRS back	2.32	0.25	[1.82 to 2.81]	p<0.001
NRS leg	1.87	0.21	[1.44 to 2.30]	p<0.001
Time	-7.68	1.37	[-10.39 to -4.98]	p<0.001
Time1	7.61	1.40	[4.84 to 10.38]	p<0.001

Table 2: Regression coefficient estimates and their 95% confidence intervals, based on the final model.

Based on these results, it can be concluded that HD-SCS is able to significantly decrease disability scores over time in patients with FBSS. Nevertheless, we should keep in mind that not for all patients data was available at each visit. Therefore, this analysis was based on all data as observed.

5.2 Sensitivity analysis

5.2.1 Missing data

Missing observations are one of the most common issues that are encountered when conducting clinical trials, however, often overlooked [15]. Let R_i be an n x 1 vector of response indicators $R_i = (R_{i1}, R_{i2}, ..., R_{in})'$ with $R_{ij} = 1$ if Y_{ij} is observed and $R_{ij} = 0$ if Y_{ij} is missing. Given R_i, the complete set of responses can be partitioned into two components Y_i^O as vector of observed responses on subject i and Y_i^M the set of responses that are missing. In 1987, Little and Rubin classified the missing data mechanisms in three distinct categories namely missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) [19]. Under the assumption of MCAR, the observed responses can be seen as a random subsample of the sampled responses. There is independence of the missing data indicator with both observed and unobserved measurements (R_i is independent of Y_i^O and Y_i^M). A more plausible assumption for clinical trials is the MAR assumption. Under this assumption, the probability of dropout depends on the observed data, but not on the unobserved data (R_i is conditionally independent of Y_i^M, given Y_i^O). Within the context of likelihood inference, MCAR and MAR are ignorable, wherefore we can ignore the missingness process and obtain valid estimates. As such, in longitudinal studies with missing data, a mixed model only requires MAR for the missing data leading to the term likelihood-based ignorable analysis. The observed data are used without removing values, nor imputing others. This strategy was applied to the model building for the primary analysis.

A wide variety of methods for handling with missing data are available. Imputation methods are commonly applied in which the missing observation is filled up with a plausible value. A commonly used technique in medicine is the 'last observation carried forward' method whereby the missing value is imputed with the last available observation. The disadvantage of this single imputation method is that it does not account for uncertainty, thereby provoking an underestimation of the standard error of the statistical point estimates [28]. A second type of imputation is multiple imputation in which the main idea is to replace every missing value by a set of M ($\geq$2) plausible values. The vector of M values is constructed based on repeated draws from the posterior predictive distribution of the unobserved values [43]. Generally, a proper imputation of Y_i^M should be randomly drawn from $f(Y_i^M|Y_i^O, X_i)$. This implicitly implies that we assume MAR during multiple imputation because the predictive distribution of the missing data, given the observed data, does not depend on the observed response pattern R_i, with $f(Y_i^M|Y_i^O, X_i, R_i) = f(Y_i^M|Y_i^O, X_i)$ [9]. These M values are presenting the uncertainty about the value, in contrast to simple imputation strategies. By the end of this step, all missing values are filled in with M values to generate M complete datasets. Standard methods are applied to analyse each dataset separately where after M inferences are then combined to withheld

one inference that is properly reflecting the sampling variability due to missing under the considered model [43]. Multiple imputation assumes MAR, however, the exact missing mechanism cannot be formally evaluated. It thus becomes clear that performing analyses on incomplete data requires untestable assumptions, so we need sensitivity analyses to understand the impact of these assumptions on inferences and conclusions from the primary analysis. Sensitivity analysis entails the creation of different models with varying assumptions and evaluating how conclusions are influenced. This rather general definition encompasses a wide variety of useful approaches [24]. In this thesis, we will focus on tipping point sensitivity analysis under MNAR.

In a tipping point analysis, the influence of missingsness is explored on the overall conclusion from the statistical inference by applying a wide spectrum of different assumptions regarding the missingness mechanisms [28]. The aim is to find the 'tipping point' in the spectrum of assumptions at which conclusions from the statistical inference will be changed [28]. Afterwards, a clinical interpretation can be given to the plausibility of the assumptions [17].

5.2.2 Sensitivity analysis of primary analysis

In this study, a substantial proportion of the data is missing. Table 3 is providing an overview of the different types of missing data. In total, 43.78% of the patients were compliant with all visits, 50.82% exhibited monotone missingness and 5.4% exhibited non-monotone missingness. Within the group with monotone missingness, a considerable amount of patients has no follow-up measurements (25.41%), 9.19% has 1 follow-up visit and 16.22% has 2 follow-up visits.

Type	M0	M1	M3	M12	number	percentage
Completers	O	O	O	O	81	43.78%
Monotone missingness	O	O	O	M	30	16.22%
	O	O	M	M	17	9.19%
	O	M	M	M	47	25.41%
Non-monotone missingness	O	O	M	O	2	1.08%
	O	M	O	O	2	1.08%
	O	M	O	M	1	0.54%
	O	M	M	O	5	2.70%

Table 3: Overview of missingness patterns in this study. Abbreviations. M: missing, O: observed.

The previously constructed mixed model is valid under MAR. However, one cannot explicitly test which mechanism is operating if only the observed data is available.

Robustness of the results under the MAR assumption was assessed by comparing the magnitude of the main effect estimated from the primary analysis to the estimates obtained from a method that assumed a MNAR mechanism. Therefore, the main inference (i.e. time effect) was explored with different assumptions about the missing data mechanism.

First, a dataset without missingness was created using multiple imputation strategies. Two approaches were consecutively performed: 1) create a dataset with only monotone missingness 2) create a dataset without missingness by regression-based imputation (PROC MI procedure). The former is achieved with a Markov Chain Monte Carlo method using a multivariate normal model [33]. The advantage of the latter is that a sequential approach with univariate models with a number of predictor variables is used. This enables first imputing data from the earliest visit, whereby the outcome can then be used as predictor for imputations at later visits [31]. The imputation model included the previous outcomes of the dependent variable combined with covariates age and pain intensity scores. Ten imputations were created for each missing value. In table 4 the main effects of the primary analysis under multiple imputation are presented. Both time effects remained significant.

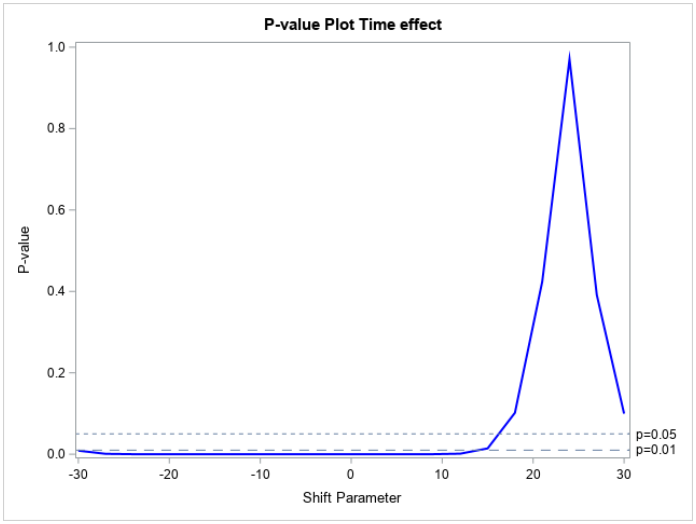
Variable	Regression estimates	SE	95% CI	p-value
Intercept	25.65	2.26	[21.19 to 30.12]	<0.001
NRS back	2.27	0.23	[1.81 to 2.73]	<0.001
NRS leg	1.83	0.20	[1.81 to 2.73]	<0.001
Time	-8.51	1.44	[-11.39 to -5.63]	<0.001
Time1	8.46	1.49	[5.49 to 11.43]	<0.001

Table 4: Regression coefficient estimates and their 95% confidence intervals, based on the final model with multiple imputation.

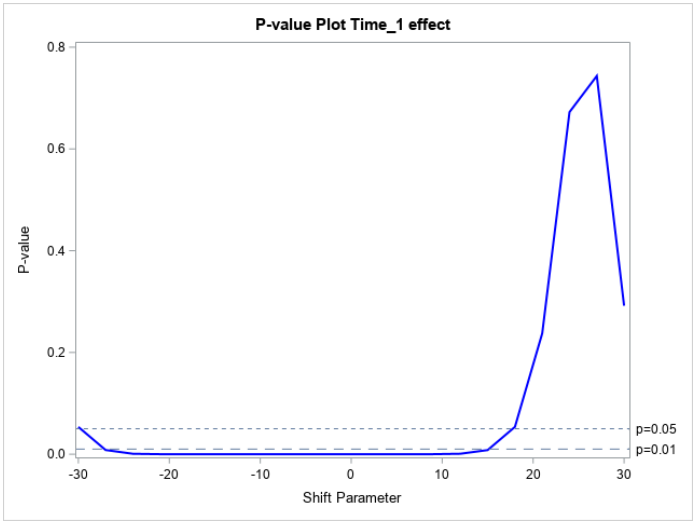
Secondly, we assumed a wide spectrum of shifts for the values that were missing, ranging from a decrease of -30 on the ODI up to + 30 at each visit. Next, the same steps as above are conducted to generate multiple imputed datasets, with a specified shift parameter that adjusts the imputed values. Thereafter, the imputed datasets were analyzed by using the same likelihood analysis as in the primary analysis (PROC MIXED procedure). Inferences are then combined for each shift parameter until a p-value of 0.05 or higher is revealed for the main study inference (PROC MI ANALYSE procedure). The basis for this analysis was the SAS macro by Yang (2013) for RCT's, available from the SAS help center.

In table 5 the results of the tipping point analysis are presented. A shift of zero corresponds to a standard MAR-based multiple imputation analysis. When ODI is shifted with a value of slightly less than 18 (meaning patients with missing values experience more disability), the conclusion becomes different from the likelihood based analysis. More

specifically, for a two-sided error level of 0.05, the tipping point for the shift parameter is 17 for time effect and 18 for time1 effect. Thus, the study conclusion under MAR is reversed when the shift parameter is 17. This means that if the shift parameter of 17 is plausible, the conclusion under MAR is questionable. Visually, the results are presented in Figure 4.



(a) Time effect



(b) Time1 effect

Figure 4: Tipping point analysis for shift parameters ranging from -30 to 30.

Shift	p-value Time	p-value Time1
-30	0.0087	0.0536
-27	0.0010	0.0082
-24	0.0001	0.0009
-21	<0.0001	0.0001
-18	<0.0001	<0.0001
-15	<0.0001	<0.0001
-12	<0.0001	<0.0001
-9	<0.0001	<0.0001
-6	<0.0001	<0.0001
-3	<0.0001	<0.0001
0	<0.0001	<0.0001
3	<0.0001	<0.0001
6	<0.0001	<0.0001
9	0.0001	0.0001
12	0.0013	0.0009
15	0.0142	0.0081
18	0.1018	0.0537
21	0.4237	0.2375
24	0.9702	0.6724
27	0.3914	0.7435
30	0.0997	0.2918

Table 5: Tipping point sensitivity analysis with p-values for time effects with shifts ranging from -30 to 30.

6 Discussion

Since 2016, a new concept in SCS has found its way in the treatment for FBSS. Spinal cord stimulation at higher current dose delivered below the sensation threshold, previously described in the literature as "high-density SCS", was demonstrated to be effective not only in pre-screened patients, but also in patients habituated to conventional SCS [41, 26]. Although the exact parameters of HD-SCS are not defined, the concept behind this new paradigm is based on the delivery of energy to neural tissue. Therefore the term "high density" became "high dose". Electrical energy may be viewed as akin to a pharmacological agent that is titrated to produce optimal pain relief [21]. The total charge delivery per unit of time (charge per pulse) seems to be a better way to describe stimulation parameters, rather than indicating a specific SCS frequency.

The primary effectiveness outcome was defined as a mean disability reduction over time by HD-SCS, measured by the ODI. Based on a longitudinal mixed model, the mean disability reduction reached a strong statistical significance over time. Moreover, the final

model contained a knot which indicated that there is a different slope from baseline to 1 month and from 1 month onward. Another approach that could have been used in this dataset was to model an exponential time curve instead of creating a knot. Approximately 20-40% of SCS patients suffer from a decline in initial effectiveness of SCS due to a central nervous system tolerance, as already reported in 1993 by LeDoux [16]. This decline is often reported for pain relief [39], however, in this study we also observed this phenomenon for ODI. Drastic improvements in ODI scores were visible up to 1 month, where after a slight decrease becomes visible. This trend was also mentioned in a study with health-related quality of life after 6 months of SCS [32]. This suggests that habituation might be an issue in SCS in general but also on the level of disability, which could potentially have a major influence on the long term clinical effects and therefore also in terms of salvage therapy and system explants [29].

The decrease in disability in FBSS patients who are treated with HD-SCS was not that surprising. In a study with multicolumn SCS, a significant decrease in ODI scores was found between baseline and 6 months of SCS [30]. In the SENZA RCT, the efficacy of high frequency SCS (another recently launched paresthesia-free SCS paradigm) was explored whereby the ODI was measured as secondary outcome variable. A significant change in ODI score was revealed after 12 months [3]. In the SENZA trial, the authors also used the ODI as an ordinal outcome measure whereby 5 distinct categories are defined based on the total ODI score. Future studies could evaluate whether the current decrease in ODI score over time is strong enough to reveal a change in category, i.e. whether the decrease is sufficiently large to reclassify patients into a lower disability category.

The International Classification of Functioning, Disability and Health (ICF) is a widely accepted framework for measuring health and disability at both individual and population levels that encompasses behavioral, physical, and integrated medical approaches [36]. Recently, the International Association for the Study of Pain Taskforce for chronic pain highlighted the importance of measuring functioning or disability based on the ICF [23] to obtain a holistic "image" of the clinical presentation of patients and to enable a better monitoring of treatment effects. In the final model of this study, pain intensity scores for low back pain and leg pain were used as predictors for disability. Several authors already focused on the association between disability and pain reportings [37, 1, 2]. In a population of patients with FBSS who are treated with HD-SCS, the degree of disability revealed a good association measures of pain intensity [7]. This study seems to confirm that pain intensity scores are important to evaluate disability in this population.

In this multicenter registry, a rather large amount of missing data was present wherefore a sensitivity analysis was performed after the primary analysis. Only one possible sensitivity analysis was performed namely the tipping-point analysis. Within this type of analysis, we

explored how severe departures from MAR must be in order to reverse conclusions from the primary analysis. In this study, a shift parameter of 17 was needed to change the main conclusions of the longitudinal mixed model. A departure of 17 is rather large, moreover, it is well above the minimal clinical important difference of the ODI which is 10 points [42]. Therefore, we can be more confident in the results obtained with statistical methods under the MAR assumptions namely the mixed model repeated measurements and multiple imputation; both pointing towards a significant time effect. Given the lack of a universally determined best MNAR method [20], one should ideally explore a variety of sensitivity analysis in order to better evaluate the consistency of results across the various assumptions that are made with different techniques.

7 Conclusion

This is the first study to report longitudinal data on disability in patients with FBSS who are treated with HD-SCS. In patients with FBSS, HD-SCS is an effective treatment option to decrease disability. Sensitivity analysis indicated that the results are maintained when the shift parameter is 17. From a clinical perspective, this shift does not seem very realistic wherefore the conclusion under MAR seems plausible.

8 Bibliography

[1] DM. Alamam, N. Moloney, A. Leaver, HI. Alsobayel, and MG. Mackey. Pain intensity and fear avoidance explain disability related to chronic low back pain in a saudi arabian population. *Spine (Phila Pa 1976)*, 44(15):E889–E898, 2019.

[2] FA. Alodaibi, JM. Fritz, A. Thackeray, SL. Koppenhaver, and JJ. Hebert. The fear avoidance model predicts short-term pain and disability following lumbar disc surgery. *PLoS One*, 13(3):e0193566, 2018.

[3] K. Amirdelfan, C. Yu, MW. Doust, BE Gliner, DM. Morgan, L. Kapural, R. Vallejo, BT. Sitzman, TL. Yearwood, R. Bundschu, T. Yang, R. Benyamin, AH. Burgher, ES. Brooks, AA. Powell, and J. Subbaroyan. Long-term quality of life improvement for chronic intractable back and leg pain patients using spinal cord stimulation: 12-month results from the senza-rct. *Qual Life Res*, 27(8):2035–2044, 2018.

[4] Z. Baber and MA. Erdek. Failed back surgery syndrome: current perspectives. *Journal of pain research*, 9:979–87, 2016.

[5] CW. Chan and P. Peng. Failed back surgery syndrome. *Pain Med*, 12(4):577–606, 2011.

[6] M. De Jaeger, L. Goudman, R. Brouns, A. De Smedt, B. Linderoth, S. Eldabe, Discover consortium, and M. Moens. The long-term response to high-dose spinal cord stimulation in patients with failed back surgery syndrome after conversion from standard spinal cord stimulation: An effectiveness and prediction study. *Neuromodulation*, Online ahead of print:–, 2020.

[7] M. De Jaeger, L. Goudman, S. Eldabe, R. Van Dongen, A. De Smedt, and M. Moens. The association between pain intensity and disability in patients with failed back surgery syndrome, treated with spinal cord stimulation. *Disabil Rehabil*, pages 1–7, 2019.

[8] J. Fairbank. Use of oswestry disability index (odi). *Spine (Phila Pa 1976)*, 20(13):1535–1537, 1995.

[9] GM. Fitzmaurice, NM. Laird, and JH. Ware. *Applied Longitudinal Analysis (Second Edition)*. John Wiley Sons, New Jersey, 2011.

[10] M. Grevitt, R. Khazim, J. Webb, R. Mulholland, and J. Shepperd. The short form-36 health survey questionnaire in spine surgery. *J Bone Joint Surg Br*, 79(1):48–52, 1997.

[11] Y. Guan. Spinal cord stimulation: neurophysiological and neurochemical mechanisms of action. *Current pain and headache reports*, 16(3):217–225, 2012.

[12] TE. Hamm-Faber, I. Gultuna, EJ. van Gorp, and H. Aukes. High-dose spinal cord stimulation for treatment of chronic low back pain and leg pain in patients with fbss, 12-month results: A prospective pilot study. *Neuromodulation*, 23(1):118–125, 2020.

[13] RG. Hazard, KF. Spratt, CM. McDonough, CM. Olson, ES. Ossen, EM. Hartmann, RJ. Carlson, and J. LaVoie. Patient-centered evaluation of outcomes from rehabilitation for chronic disabling spinal disorders: the impact of personal goal achievement on patient satisfaction. *Spine J*, 12(12):1132–1137, 2012.

[14] DJHA. Henssen, N. Scheepers, E. Kurt, I. Arnts, M. Steegers, K. Vissers, R. van Dongen, and Y. Engels. Patients' expectations on spinal cord stimulation for failed back surgery syndrome: A qualitative exploration. *Pain Pract*, 18(4):452–462, 2018.

[15] MA. Klebanoff and SR. Cole. Use of multiple imputation in the epidemiologic literature. *Am J Epidemiol*, 168(4):355–357, 2008.

[16] MS. LeDoux and KH. Langford. Spinal cord stimulation for the failed back syndrome. *Spine*, 18(2):191–4, 1993.

[17] B. Leurent, M. Gomes, R. Faria, S. Morris, R. Grieve, and JR. Carpenter. Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: A tutorial. *Pharmacoeconomics*, 36(8):889–901, 2018.

[18] B. Linderoth and RD. Foreman. Conventional and novel spinal stimulation algorithms: Hypothetical mechanisms of action and comments on outcomes. *Neuromodulation*, 20(6):525–33, 2017.

[19] R. Little and D. Rubin. *Statistical Analysis With Missing Data*. John Wiley Sons, New York, 1987.

[20] CH. Mallinckrodt, Q. Lin, and M. Molenberghs. A structured framework for assessing sensitivity to missing data assumptions in longitudinal clinical trials. *Pharm Stat*, 12(1):1–6, 2013.

[21] JP. Miller, S. Eldabe, E. Buchser, LM. Johanek, Y. Guan, and B. Linderoth. Parameters of spinal cord stimulation and their role in electrical charge delivery: A review. *Neuromodulation*, 19(4):373–84, 2016.

[22] G. Molenberghs and G. Verbeke. *Models for Discrete Longitudinal Data*. Springer, New York, 2005.

[23] B. Nugraha, C. Gutenbrunner, A. Barke, M. Karst, J. Schiller, P. Schäfer, S. Falter, B. Korwisi, W. Rief, RD. Treede, and IASP Taskforce for the Classification of Chronic Pain. The iasp classification of chronic pain for icd-11: functioning properties of chronic pain. *Pain*, 160(1):88–94, 2019.

[24] National Research Council (US) Panel on Handling Missing Data in Clinical Trials. *The Prevention and Treatment of Missing Data in Clinical Trials: 5, Principles and Methods of Sensitivity Analyses*. Washington (DC): National Academies Press (US). 2010.

[25] JE. Pope and M. Fishman. Redefining success: Longitudinal patient reported outcome measures and the importance of psychometric testing for optimization in neuromodulation. *Neuromodulation*, 22(1):119–20, 2019.

[26] DA. Provenzano, J. Rebman, C. Kuhel, H. Trenz, and J. Kilgore. The efficacy of high-density spinal cord stimulation among trial, implant, and conversion patients: A retrospective case series. *Neuromodulation*, 20(7):654–60, 2017.

[27] A. Ramasamy, ML. Martin, SI. Blum, H. Liedgens, C. Argoff, R. Freynhagen, M. Wallace, KP. McCarrier, DM. Bushnell, NV. Hatley, and DL. Patrick. Assessment of patient-reported outcome instruments to assess chronic low back pain. *Pain Med*, 18(6):1098–1110, 2017.

[28] B. Ratitch, M. O'Kelly, and R. Tosiello. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharm Stat*, 12(6):337–47, 2013.

[29] RD. Reddy, R. Moheimani, GG. Yu, and KV. Chakravarthy. A review of clinical data on salvage therapy in spinal cord stimulation. *Neuromodulation*, pages –, 2019.

[30] P. Rigoard, S. Basu, M. Desai, R. Taylor, L. Annemans, Y. Tan, MJ. Johnson, C. Van den Abeele, R. North, and PROMISE Study Group. Multicolumn spinal cord stimulation for predominant back pain in failed back surgery syndrome patients: A multicenter randomized controlled trial. *Pain*, 160(6):1410–1420, 2019.

[31] DB. Rubin. *Multiple imputation for nonresponse in surveys*. John Wiley and Sons, New York, 1987.

[32] L. Scalone, F. Zucco, A. Lavano, A. Costantini, M. De Rose, P. Poli, G. Fortini, L. Demartini, E. De Simone, V. Menardo, M. Meglio, P. Cozzolino, PA. Cortesi, and LG. Mantovani. Benefits in pain perception, ability function and health-related quality of life in patients with failed back surgery syndrome undergoing spinal cord stimulation in a clinical practice setting. *Health and quality of life outcomes*, 16(1):68, 2018.

[33] JL. Schafer. *Analysis of incomplete multivariate data*. Chapman and Hall, New York, 1997.

[34] A. Sebaaly, MJ. Lahoud, M. Rizkallah, G. Kreichati, and K. Kharrat. Etiology, evaluation, and treatment of failed back surgery syndrome. *Asian spine journal*, 12(3):574–585, 2018.

[35] CN. Shealy, JT. Mortimer, and JB. Reswick. Electrical inhibition of pain by stimulation of the dorsal columns: preliminary clinical report. *Anesth Analg*, 46(4):489–91, 1967.

[36] G. Stucki, A. Cieza, T. Ewert, N. Kostanjsek, S. Chatterji, and TB. Ustün. Application of the international classification of functioning, disability and health (icf) in clinical practice. *Disabil Rehabil*, 24(5):281–282, 2002.

[37] S. Tarvonen-Schroder, A. Kaljonen, and K. Laimi. Comparing functioning in spinal cord injury and in chronic spinal pain with two icf-based instruments: Whodas 2.0 and the who minimal generic data set covering functioning and health. *Clin Rehabil*, 33(7):1241–1251, 2019.

[38] RS. Taylor, MJ. Desai, P. Rigoard, and RJ. Taylor. 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*, 14(6):489–505, 2014.

[39] SJ. Thomson, D. Kruglov, and RV. Duarte. A spinal cord stimulation service review from a single centre using a single manufacturer over a 7.5 year follow-up period. *Neuromodulation*, 20(6):589–99, 2017.

[40] G. Verbeke and G. Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer, New York, 2009.

[41] F. Wille, JS. Breel, EW. Bakker, and MW. Hollmann. Altering conventional to high density spinal cord stimulation: An energy dose-response relationship in neuropathic pain therapy. *Neuromodulation*, 20(1):71–80, 2017.

[42] A. Wright, J. Hannon, EJ. Hegedus, and AE. Kavchak. Clinimetrics corner: a closer look at the minimal clinically important difference (mcid). *J Man Manip Ther*, 20(3):160–166, 2012.

[43] X. Yan, S. Lee, and N. Li. Missing data handling methods in medical device clinical trials. *Journal of Biopharmaceutical Statistics*, 19(6):1085–1098, 2009.

9 Appendix

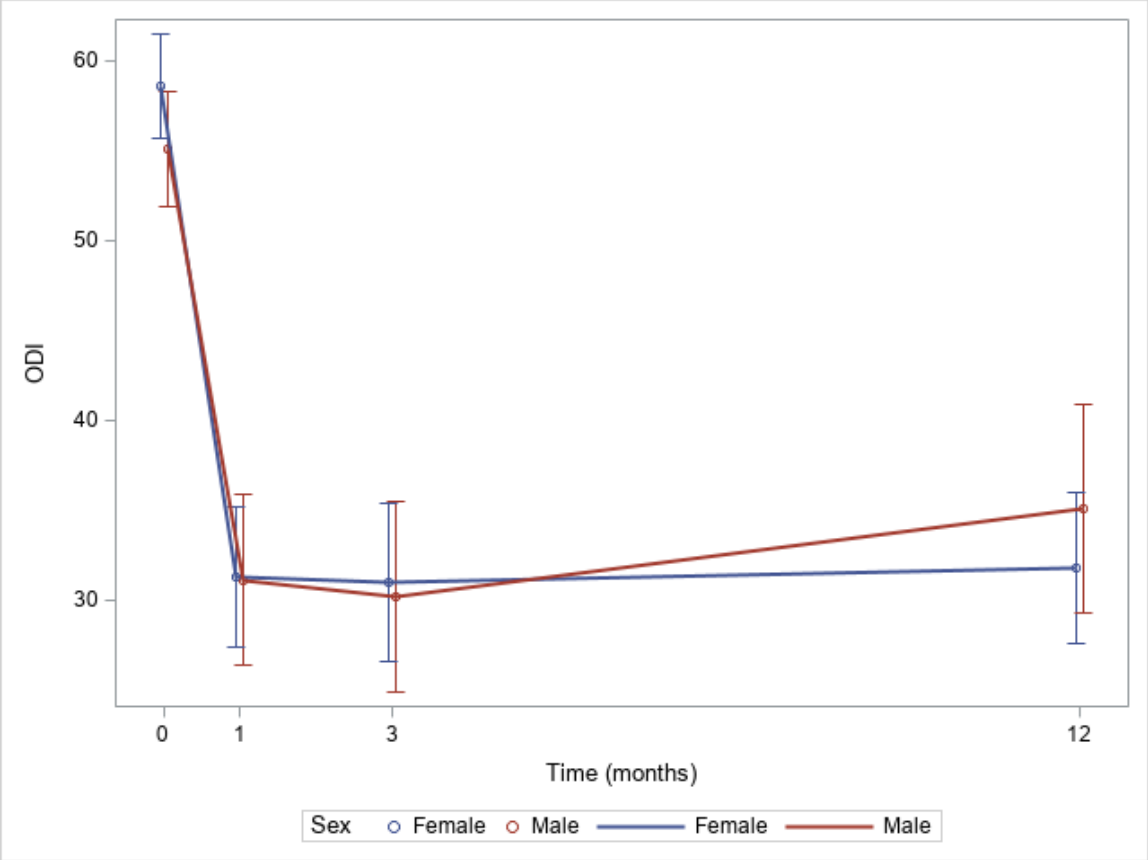


Figure A1: *Mean profile plots (with 95% confidence intervals) for ODI in function of visit according to sex.*

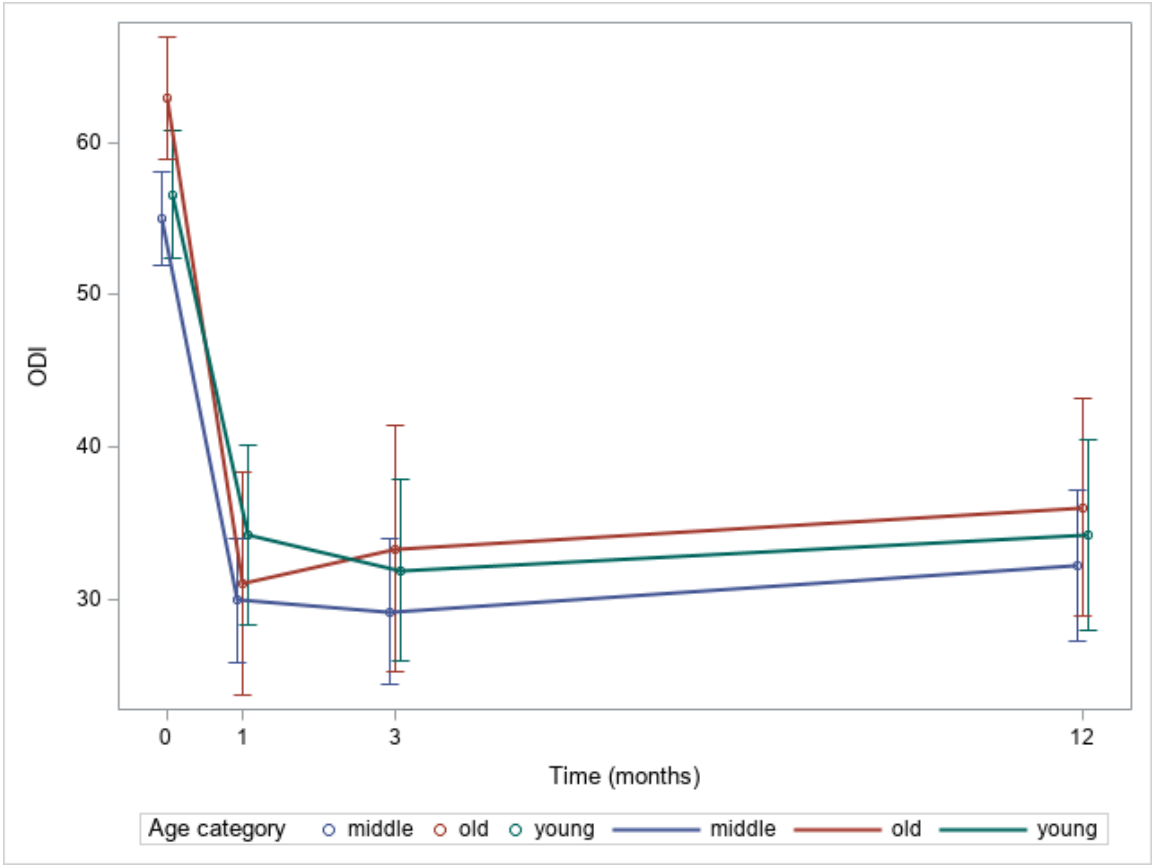


Figure A2: *Mean profile plots (with 95% confidence intervals) for ODI in function of visit according to age category.*

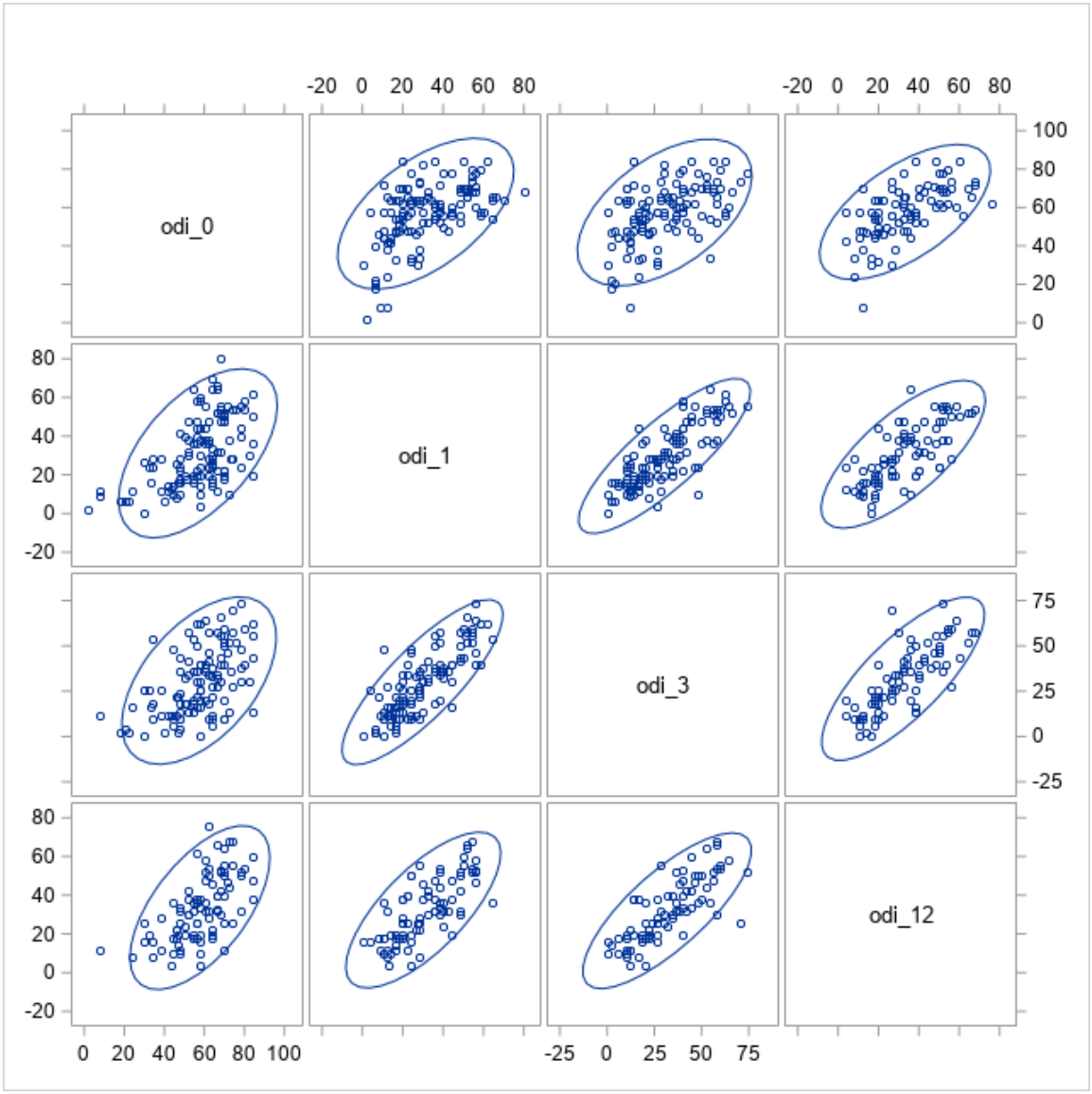


Figure A3: *Scatterplot matrix of the correlations of ODI scores between visits.*