



Universiteit  
Leiden  
The Netherlands

## Outcome after anterior cervical discectomy: from inferential statistics to Machine Learning

Goedmakers, C.M.W.

### Citation

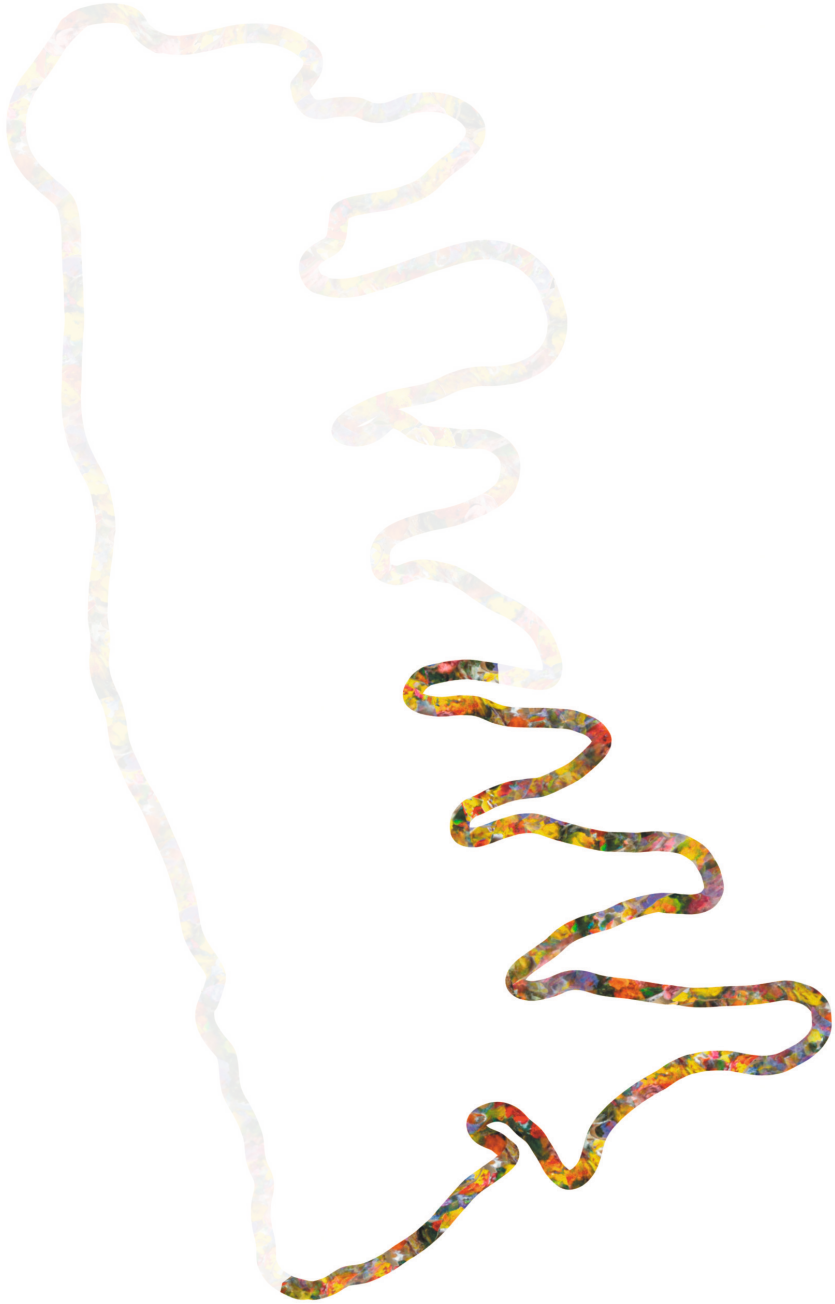
Goedmakers, C. M. W. (2023, December 20). *Outcome after anterior cervical discectomy: from inferential statistics to Machine Learning*. Retrieved from <https://hdl.handle.net/1887/3674247>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3674247>

**Note:** To cite this publication please use the final published version (if applicable).



# The Impact of Mental Health on Outcome After Anterior Cervical Discectomy

Cohort study Assessing the Influence of Mental Health Using Predictive Modelling

---

Goedmakers CMW<sup>1</sup>, Ilse van Beelen<sup>2</sup>, Floor Komen<sup>2</sup>,  
Erik W. van Zwet<sup>2</sup>, Wilco C. Peul<sup>1</sup>, Mark P. Arts<sup>1</sup>,  
Carmen L.A. Vleggeert-Lankamp<sup>1,2</sup>

*Acta Neurochir (Wien)*. 2022 Nov;164(11):3035-3046.

<sup>1</sup>Neurosurgical Center Holland, Leiden University Medical Center & Haaglanden MC & Haga Teaching Hospital, <sup>2</sup>Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands, <sup>3</sup> Department of Neurosurgery, Spaarne Gasthuis Haarlem/Hoofddorp, the Netherlands

## Abstract

### Background

Depression and anxiety are common mental disorders among patients with chronic pain. It is hypothesized that patients suffering from these disorders benefit less from cervical spine surgery than mentally healthy patients. Therefore this study aimed to quantify the effect of mental health status on functional outcome after anterior cervical discectomy in a post-hoc analysis on RCT data.

### Methods

108 patients from the NECK-trial, with radiculopathy due to a one level herniated disc, underwent anterior cervical discectomy and were included into this analysis. Functional outcome was quantified using the Neck Disability Index (NDI) and mental health status was measured using the Hospital Anxiety and Depression Score (HADS) questionnaire. NDI differences were assessed using Generalized Estimated Equations (GEE), crude means, a predictive Linear Mixed Model (LMM) using baseline scores and over time with an explanatory LMM.

### Results

At baseline, 24% and 32% of patients were respectively depressed and anxious, and had statistically significant and clinically relevant higher NDI scores during follow-up. However, in those patients in which the HADS returned to normal during follow up, NDI values decreased comparably to the non-depression or -anxiety cases. Those patients that demonstrated persisting high HADS values, had convincingly worse NDI scores. A predictive Linear Mixed Model (LMM) showed that combining baseline NDI and HADS scores was highly predictive of NDI during follow-up. The R-shiny application enabled the effective, visual communication of results from the predictive LMM.

### Conclusion

This study shows that mental health status and disability are strongly associated and provides insight into the size of the effect, as well as a way to use this relation to improve preoperative patient counselling. These findings give rise to the suggestion that incorporating mental health screening in the pre-operative assessment of patients could help to adequately manage patients' expectations for functional recovery.

## Introduction

Depression and anxiety are common psychological disorders in patients with pain and chronic diseases. A review in JAMA internal medicine reports a 56% mean prevalence of major depression in patients having pain in orthopaedic or rheumatology disease [1]. At the same time, depressive symptoms are known to influence clinical outcome in patients being treated for pain. Patients suffering from depressive symptoms report more pain, more intense pain, more amplification of pain symptoms, and longer duration of pain [2-5]. Additionally, patients with both conditions have a lower self-perceived recovery rate and are more likely to report persistent pain [5-8].

In a systematic review on the relation between psychological disorders and spine surgery, it was concluded that this group of patients suffers from higher rates of spinal pain, postoperative complications and worsened functional outcomes [9]. Specifically, in patients with lumbar radiculopathy it was demonstrated that better mental health at baseline was significantly associated with lower disability after surgery [10]. In the cervical spine however, the relation between mental health and functional outcome after surgery is less well investigated. One study showed statistically significant improvement in postoperative neck pain after one year in patients who received treatment for their anxiety compared those who had not [11]. However, longer follow-up on these patients with treated depression demonstrated no significant difference in objective or subjective outcomes up to 24 months after surgery [12]. Another study that described the impact of preoperative depression on outcome after posterior cervical fusion and found that depressed patients reported less improvement in postoperative quality of life [13], but this could not be confirmed in a later study [14]. The true size of the effect of mental health on functional outcome after cervical spine surgery thus remains to be unknown. Ultimately, not just the effect size of mental health on functional outcome is of interest, but also how the association can be used to effectively counsel patients preoperatively.

Therefore, in this study the relation between mental health and functional outcome after anterior cervical spine surgery were prospectively studied based on RCT data. Firstly, the effect size of mental health status on functional outcome was quantified. Secondly, a prediction model was developed and implemented in an application to improve preoperative patient counselling in clinical practice.

## Materials and Methods

### Design

In this study a post-hoc analysis was performed on data collected as part of the NECK-trial, a prospective, double-blinded multicentre randomized controlled trial (RCT), conducted among patients with cervical radiculopathy due to single-level disc herniation. Patients were randomly assigned into three groups: anterior cervical discectomy with arthroplasty (ACDA), anterior cervical discectomy and fusion (ACDF), and anterior cervical discectomy alone (ACD). Details about the protocol, inclusion criteria, sample size calculations, methods and outcomes of this trial have been previously published [15, 16]. The trial showed small, non-significant and not clinically relevant

differences in clinical outcome between the three treatment groups after two years. Therefore, all patients from the NECK-trial were analysed as one cohort in this study.

### **Outcome measures**

Data was prospectively collected. The primary clinical outcome measures used in the NECK-trial was Neck Disability Index (NDI). To assess mental health status the Hospital Anxiety and Depression Scale (HADS) was used. Data on the HADS scores have not previously been published.

The HADS is a patient-reported questionnaire to screen for generalized anxiety disorder (GAD) and depression. The questionnaire consists of 14 items, half of these focus on depression and the other half on anxiety. The HADS score classifies patients into three categories; cases (11-21 points), doubtful cases (8-10 points) or non-cases (0-7 points) for GAD and depression separately [17, 18]. In addition to the inclusion criteria used in the NECK-trial, patients needed to have baseline HADS measurement in order to be included into this analysis.

Neck Disability Index (NDI) was used to measure functional outcome. The NDI is a 10-item questionnaire on three different aspects; pain intensity, daily work related activities and non-work related activities. The total score ranges from 0 (best score) to 50 (worst score) and was converted to a 100 point scale. The NDI is a modification of the Oswestry Low Back Pain Questionnaire and has been shown to be reliable and valid for patients with cervical pathology [19-21]. Patients were asked to fill out both HADS and NDI questionnaires at baseline, one and two years after surgery.

### **Statistical analysis**

The statistical analysis is performed with R version 3.6.0 and Rstudio version 1.2.1335. All code is shared in an online, open data repository (Appendix A).

#### ***Statistics using HADS baseline score***

Patient demographics were analysed grouped upon baseline HADS scores (cases, doubtful cases and non-cases), and tested using the chi-squared test for categorical values, ANOVA tests for parametric numerical variables and Kruskal Wallis tests for nonparametric numerical variables. Data was analysed separately for HADS anxiety and HADS depression.

To study how the NDI scores developed over time for each baseline HADS group, Generalized Estimating Equations (GEE) were used. Analysing repeated measurements using GEE allows for estimates of the outcome, based on variation within individuals. The multiple measurements can control for the time-invariant, and unobservable differences between individuals. In this model the follow-up moment, the HADS category and the individual patient numbers were used to explain the dependent variable; NDI.

To study how baseline HADS and baseline NDI scores can be used to predict NDI scores one and two years after surgery a predictive Linear Mixed Model (LMM) was developed. NDI at baseline and HADS at baseline are centred at its mean, which improves interpretability of the intercept. Part of the within-group variance can be accounted for by adding a random intercept in a LMM.

The Intra Class Correlation (ICC) was calculated to quantify the amount of within-group variance that the random intercept can explain.

The predictive ability of the model is tested using two different methods; predictions for four specific patients and cross validation (CV). An R-shiny application will be developed to implement and visualize the results from the predictive LMM.

### ***Statistics using HADS over time***

As HADS scores can change over time, the relation between change in HADS was studied in relation to change in NDI scores. Firstly, the Pearson correlation coefficient was calculated between decrease in HADS depression, HADS anxiety and NDI with corresponding p-values. Secondly, patients were divided into four groups, based on their change in HADS scores (delta HADS) to study how NDI change related to HADS group change. Patients were categorized into either; 1. (Doubtful) case at baseline, no (doubtful) case after two years; 2. (Doubtful) case at baseline, (doubtful) case after two years; 3. No (doubtful) case at baseline, (doubtful) case after two years; 4. No (doubtful) case at baseline, no (doubtful) case after two years. There was no distinction made between anxiety and depression in delta HADS groups.

Additionally, in order to study HADS as a continuous variable, an explanatory LMM was used to analyse the dependent variable NDI over time, in relation to total HADS and follow-up time as independent variables. HADS is centred at its mean, for the same reasons as mentioned for the predictive LMM.

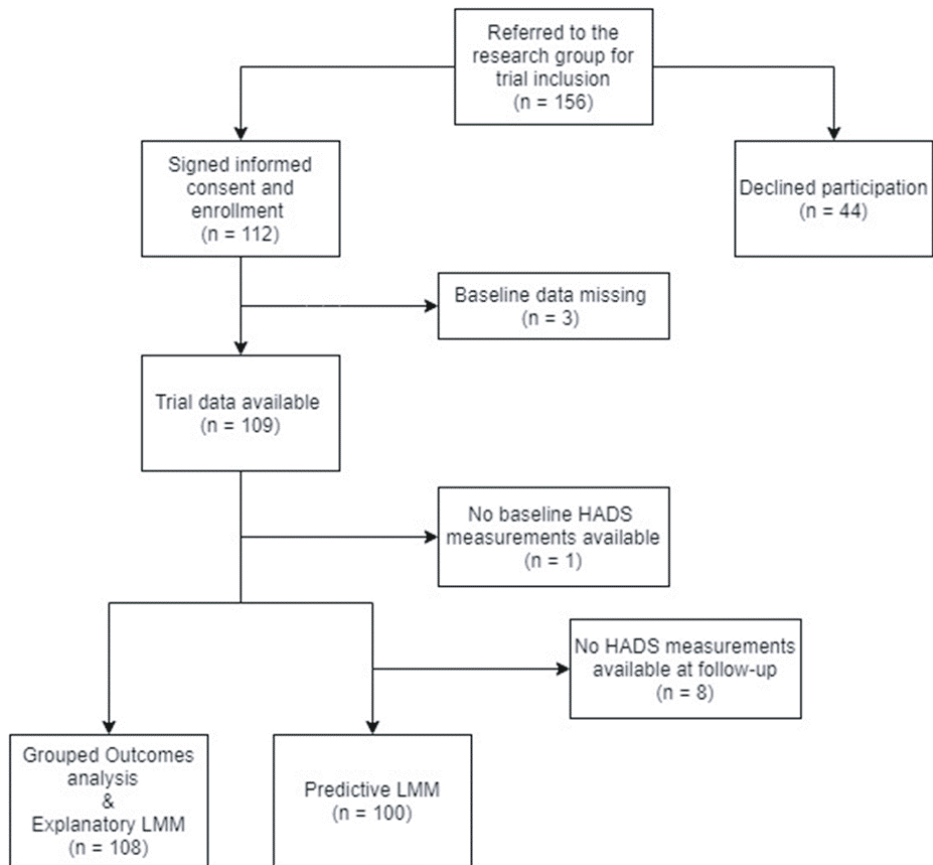
## **Results**

### **1. HADS at baseline**

#### **1.1 Baseline characteristics**

108 patients were included in this analysis; patients needed to have at least completed NDI and HADS measurements at baseline (Figure 1). Patients were on average  $46.8 \pm 7.9$  years old, 52% was female, BMI was  $26.6 \pm 4.3$ , NDI score was  $44.4 \pm 15.4$  and the median duration of complaints was 26.0 (IQR 39) weeks. At baseline 12 patients (11%) were classified as 'depression cases' and 14 patients (13%) were classified as 'GAD cases'. 10 of those patients were classified as both depression and GAD cases.

NDI at baseline was significantly higher, indicating higher disability, for both the depression and the anxiety cases as compared to the doubtful and the non-cases ( $p < 0.001$ ; Table 1). There were no other statistically significant differences between cases, doubtful cases and non-cases at baseline when grouping on either baseline HADS depression or anxiety scores.



**Figure 1.** Flow chart illustrating the inclusion process

### 1.2 Evaluation of NDI in patients grouped by baseline HADS score

Mean values for NDI decreased significantly from 41-47 points at baseline to 18-21 points at 1 year and 19-20 points at 2 year follow up in all three treatment groups ( $p < 0.001$ ), without significant differences between the three treatment arms (ACD, ACDF or ACDA) [16].

Grouped by the HADS depression score at baseline, depression cases report a marginal mean NDI that was more than doubled at baseline (28.0 points higher) and two years after surgery (27.5 points higher) when compared to non-cases (Table 2, Figure 2). The difference is not only statically significant ( $p < 0.001$ ) but more importantly clinically relevant, as the Minimal Clinically Important Difference (MCID) for NDI is estimated at 20 points [22-25]. Likewise, depression cases showed comparably higher scores in comparison to doubtful depression cases at all timepoints, while there were small differences in NDI between doubtful and non-cases.



Table 1. Patient characteristics per baseline HADS group and total.

	Depression		Anxiety		Total
	Non-cases	Doubtful cases	Non-cases	Doubtful cases	
<b>Number (n)</b>	82	14	73	21	108
<b>Age (yr)</b>	46.6 ± 8.4	45.6 ± 4.4	46.6 ± 8.2	47.6 ± 6.5	46.8 ± 7.9
<b>Sex (F/M)</b>	43/39	5/9	38/35	10/11	56/52
<b>BMI</b>	26.6 ± 4.3	26.6 ± 3.9	26.7 ± 4.3	25.6 ± 4.7	26.6 ± 4.3
<b>NDI</b>	40.7 ± 13.3*	45.3 ± 11.8*	40.1 ± 13.5*	45.9 ± 9.9*	44.4 ± 15.4
<b>Duration complaints (weeks)</b>	26.0 (39)	32.5 (36)	26.0 (39)	26.0 (35)	26.0 (39)

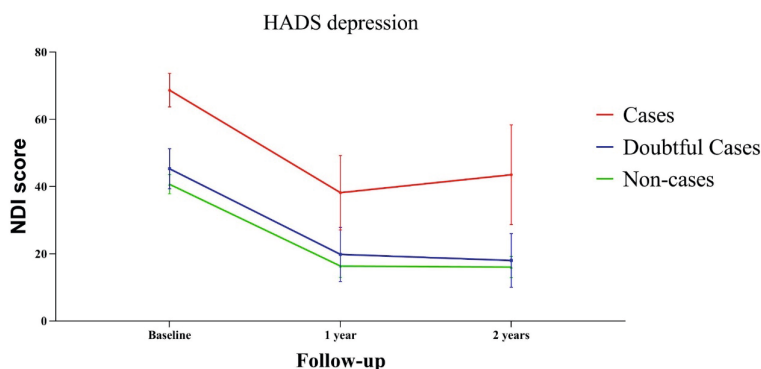
\* Indicates a significant difference between the groups. All other characteristics were similar between groups without reaching statistical significance. Parametric numerical data was represented by mean value ± standard deviation (SD) and nonparametric data as median (interquartile range).

**Table 2.** The difference in estimated marginal mean NDI between HADS depression groups calculated using Generalized Estimated Equations.

Follow-up	Comparison between depression status		Mean NDI difference	SE	Lower bound 95% CI	Upper bound 95% CI	P-value
<b>Baseline</b>	Cases	Non-cases	28.0	2.9	22.2	33.8	< 0.001
	Cases	Doubtful cases	23.4	4.0	15.6	31.1	< 0.001
	Doubtful cases	Non-cases	4.6	3.4	-2.0	11.2	0.170
<b>1 year post-operative</b>	Cases	Non-cases	21.8	5.9	10.3	33.4	< 0.001
	Cases	Doubtful cases	18.4	7.0	4.7	32.0	0.009
	Doubtful cases	Non-cases	3.5	4.4	-5.2	12.2	0.434
<b>2 years post-operative</b>	Cases	Non-cases	27.5	7.7	12.3	42.7	< 0.001
	Cases	Doubtful cases	25.5	8.6	8.7	42.4	0.003
	Doubtful cases	Non-cases	2.0	4.4	-6.6	10.6	0.651

\* Groups based on HADS anxiety scores at baseline

\*\* HADS = Hospital Anxiety and Depression Scale, SE = Standard Error, NDI = Neck Disability Index, CI = Confidence Interval



\* Groups based on the HADS depression scores at baseline

**Figure 2.** NDI during follow-up for each HADS depression group. HADS depression groups are based on the baseline HADS scores reported.

\* Groups based on HADS depression scores at baseline.

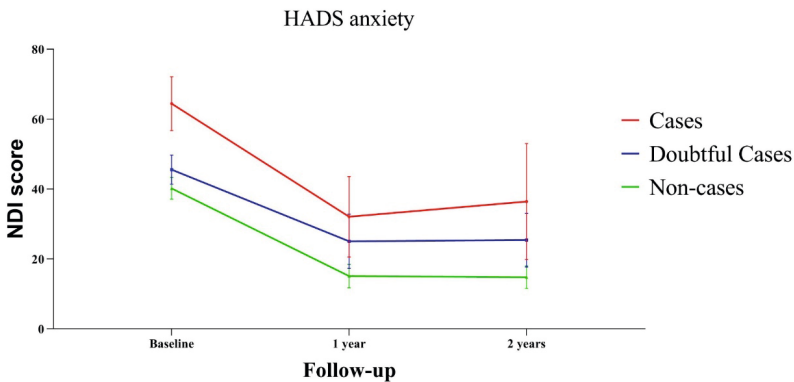
Grouped according to baseline HADS anxiety score, anxiety cases report a 24.2, 17.0 and 21.6 points higher NDI score when compared to non-cases, respectively at baseline, one year after surgery and two years after surgery (Table 3, Figure 3). The difference is statistically significant ( $p < 0.001$ ,  $p = 0.006$ ,  $p = 0.012$ ) and exceeds the MCID. At baseline the difference between the anxiety cases and doubtful cases is 18.9 ( $p < 0.001$ ), however one and two years post-operatively the differences between cases and doubtful are similar to those between the doubtful cases and non-cases (Table 3, Figure 3).

**Table 3.** The difference in estimated marginal mean NDI between HADS anxiety groups calculated using Generalized Estimated Equations.

Follow-up	Comparison between anxiety status		Mean NDI difference	SE	Lower bound 95% CI	Upper bound 95% CI	P-value
Baseline	Cases	Non-cases	24.2	4.2	15.9	32.5	< 0.001
	Cases	Doubtful cases	18.9	4.5	10.1	27.6	< 0.001
	Doubtful cases	Non-cases	5.4	2.6	0.2	10.6	0.042
1 year post-operative	Cases	Non-cases	17.0	6.1	5.0	28.9	0.006
	Cases	Doubtful cases	7.0	7.1	- 6.9	20.9	0.323
	Doubtful cases	Non-cases	10.0	4.3	1.5	18.4	0.021
2 years post-operative	Cases	Non-cases	21.6	8.6	4.7	38.5	0.012
	Cases	Doubtful cases	11.0	9.3	- 7.2	29.3	0.237
	Doubtful cases	Non-cases	10.6	4.3	2.3	19.0	0.013

\* Groups based on HADS anxiety scores at baseline

\*\* HADS = Hospital Anxiety and Depression Scale, SE = Standard Error, NDI = Neck Disability Index, CI = Confidence Interval



\* Groups based on the HADS anxiety scores at baseline

**Figure 3.** NDI during follow-up for each HADS anxiety group. HADS anxiety groups are based on the baseline HADS scores reported.

\* Groups based on HADS anxiety scores at baseline

### 1.3 Predicting NDI based on baseline HADS score

A predictive LMM was used to predict the NDI values after 52 and 104 weeks, based on the baseline NDI and HADS values. The estimates of the beta coefficients were respectively 0.37 ( $p < 0.001$ ) and 0.73 ( $p < 0.001$ ) (Table 4). The beta coefficient calculated for time in weeks, was not significant ( $p = 0.75$ ) most likely due to the small average differences in NDI scores between the two time points.

**Table 4.** Coefficients of the predictive Linear Mixed Effects Model with its corresponding standard error (SE), degrees of freedom (DF), t-value and p-value.

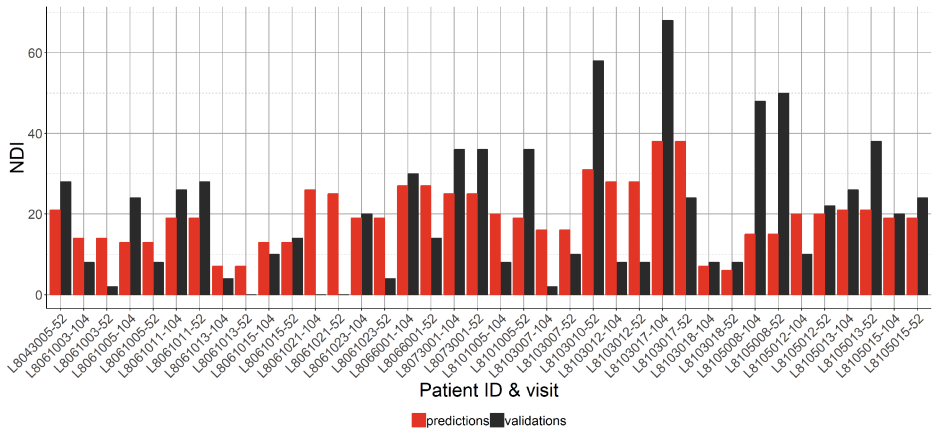
	<b>Coefficients</b>	<b>SE</b>	<b>DF</b>	<b>t-value</b>	<b>p-value</b>
(Intercept)	18.98	2.43	161.49	7.82	<0.001
Deviation NDI at baseline	0.37	0.11	99.89	3.29	<0.001
Deviation HADS at baseline	0.73	0.24	98.48	3.07	<0.001
Time in weeks	0.0081	0.026	91.89	0.32	0.75
Time with symptoms	-0.18	0.27	16.90	-0.65	0.5229
VAS Neck pain	0.06	0.07	96.42	0.92	0.3590
VAS Arm pain	0.06	0.053	96.73	1.15	0.2550
Osteofyte / Spondylosis	-5.93	4.36	56.35	-1.36	0.1789
Positive family history for neck problems	-2.81	6.35	96.72	-0.44	0.6591
Age	0.14	0.18	102.44	0.78	0.4364
Gender	2.53	2.83	98.85	0.89	0.3744
Smoking	0.55	2.90	99.70	0.19	0.8494
Alcohol use	1.19	3.08	98.96	0.39	0.6999
BMI	-0.40	0.33	94.61	-1.23	0.2223
Disc height at index	-1.35	1.12	84.77	-1.21	0.2312

*Given the mean NDI of 44 and the mean HADS of 11, if a patient has an NDI of 54 (deviation of 10 with the mean NDI) and a HADS of 10 (deviation of -1 with the mean HADS) at baseline, the NDI after two years will be predicted as:  $18.98 + 0.37 \cdot 10 + 0.73 \cdot -1 + 0.0081 \cdot 104 = 22.79$ .*

The RMSE  $\pm$  SD is after a 5-fold cross validation (CV) calculated to be  $14.5 \pm 1.7$  after 52 weeks and  $15.8 \pm 1.6$  after 104 weeks and did not show a clear pattern of model under or over-prediction (Figure 4). Comparing individual predictions shows that predictions for patients that have much higher than average NDI, at either 52 weeks or 104 weeks, are not within the confidence interval of the predicted value (Figure 5).

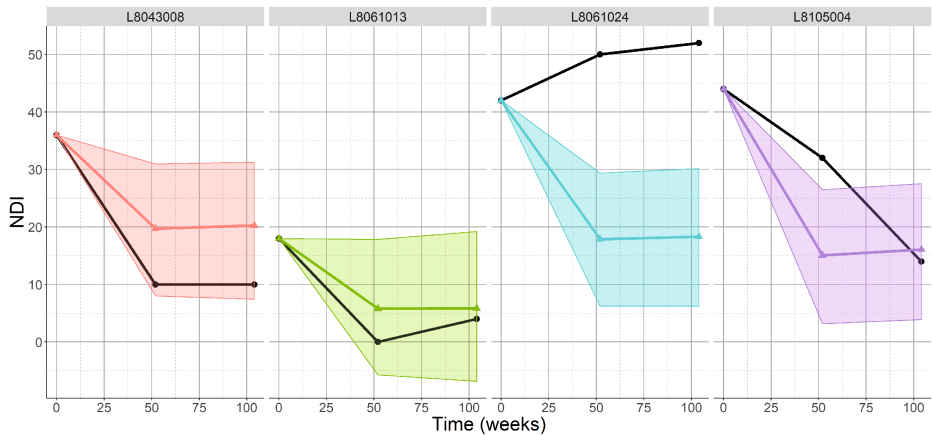
The intra-class correlation (ICC) was 0.75. NDI and HADS at baseline are therefore highly predictive for NDI at 52 and 104 weeks. To investigate whether the remaining 0.25 variance could be further explained, other covariates were added to the predictive LMM. Weight, beta coefficient -0.19 ( $p = 0.0313$ ) and height, beta coefficient -0.32 ( $p = 0.0193$ ) were both statistically significant (Table 4), while BMI was not ( $p = 0.2223$ ). Time with symptoms; VAS neck pain; VAS arm pain; osteophyte / spondylosis; positive family history; age; gender; smoking; alcohol use and disc height were also not statistically significant.

In order to visualize the results of the predictive LMM, an R-shiny app was developed. The app allows the physician to communicate the results of the LMM visually to the patient using a dynamic graph (Figure 6).



**Figure 4.** 5-fold cross validation

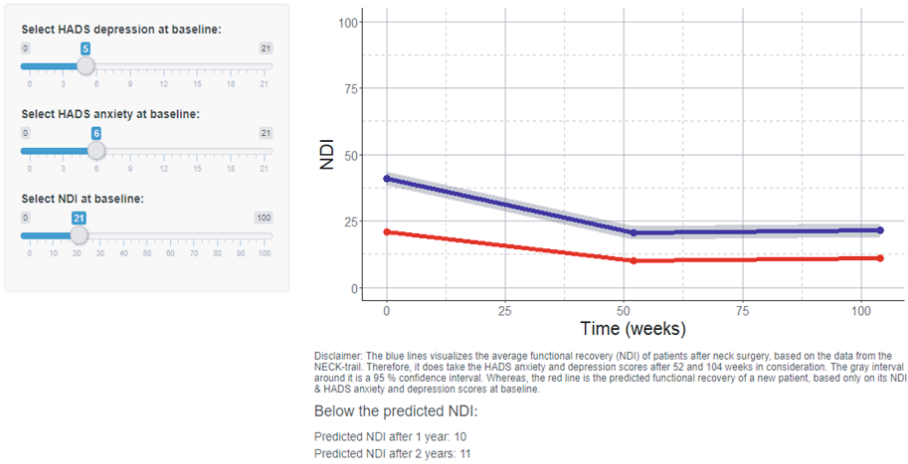
5-fold cross validation (CV) with the predictions (in red) and actual NDI (in dark gray) of one fold. The x-axis visualizes the specific anonymized patient ID combined with a visit (52 or 104 weeks). For CV the data is randomly divided into 5 parts, where 4 parts will function as the training data and the 5th part as the test data. The procedure is repeated 5 times. The lower the Root Mean Square Error (RMSE), the better the predictive ability of the model, and the RMSE can also be interpreted as the average amount of NDI points that the model predicts less or more than the actual NDI.



**Figure 5.**

Predicted NDI values on week 52 and 104 and their confidence intervals for four individual patients, shown in color. The black lines are the true NDI values. In the patient specific predictions the training data is created by removing one patient from the dataset. The predictive LMM is trained on this training data and predicts for the left out patient. This procedure is repeated four times, for four different patients.

### Functional outcome after ACD, ACDF or ACDA



### Functional outcome after ACD, ACDF or ACDA

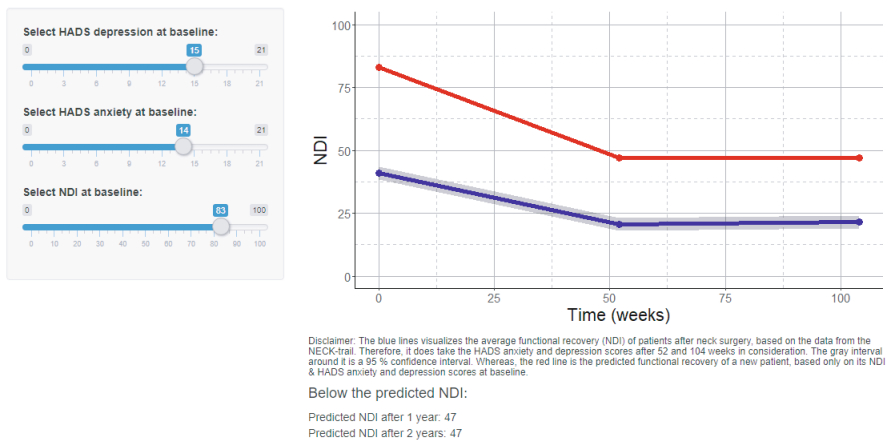


Figure 6. R shiny

Screenshots of the R shiny application illustrating how it functions. On the left the input can be given and on the right the predicted NDIs will be visualized in the graph (red line). On the left side the adjustable baseline measurements for NDI, HADS anxiety and HADS depression. On the right the graph visualizing (red line) the predicted NDI on the y-axis during the follow-up moment in weeks on the x-axis, and the marginal mean NDI (blue line) over time with a 95% confidence interval (in grey). Beneath the graph, the numerical values appear for the predicted NDI at week 52 and 104.

## 2. HADS during follow-up

### 2.1 Evaluation of NDI in patients grouped by delta HADS score

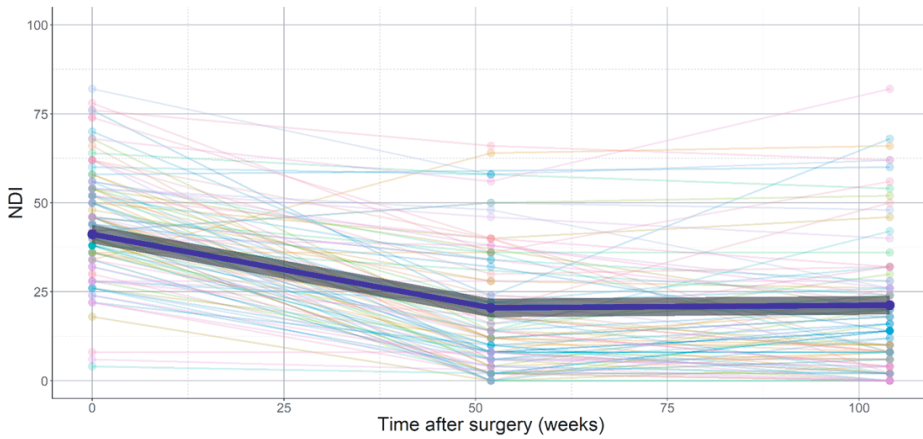
Decrease in HADS depression score was correlated to a decrease in NDI, and likewise the decrease in HADS depression was correlated to a decrease in HADS anxiety (Pearson correlation coefficient 0.544 and 0.552, respectively ( $p < 0.001$ )) (Table 5, Appendix D)). Furthermore, the decrease in NDI during the first year was comparable in the three groups (cases, doubtful cases and non-cases), and the value of NDI remained stable after one year in all groups (based on baseline HADS scores) (Figure 2, 3, 7). However, HADS scores were measured again after one and after two years and changed during follow-up in some patients. To evaluate the effect on NDI, patients were additionally differentiated based on their change in HADS (delta HADS) into; no (doubtful) case at baseline, (doubtful) case after two years ( $n = 7$ ); no (doubtful) case at baseline, no (doubtful) case after two years ( $n = 59$ ); (doubtful) case at baseline, no (doubtful) case after two years ( $n = 22$ ) or (doubtful) case at baseline, (doubtful) case after two years ( $n = 15$ ). Remarkably, NDI changed in the same direction as the HADS score changes (Figure 8, Appendix E). Patients with increasing HADS scores, report higher NDI scores, an interaction that can also be seen with change in the opposite direction; with a decreasing HADS and lower NDI scores (Figure 8, Appendix E).

**Table 5.** Correlations between decrease in HADS depression, HADS anxiety and NDI from baseline to two years after surgery.

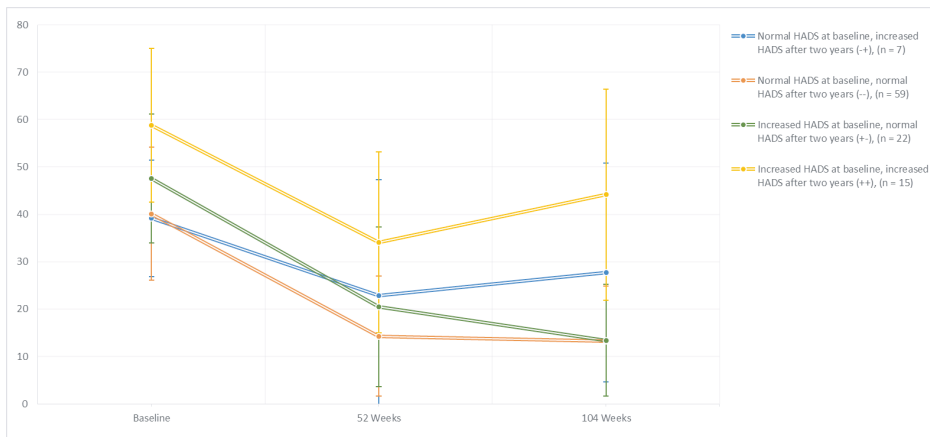
		Decrease in NDI	Decrease in HADS depression score	Decrease in HADS anxiety score
<b>Decrease in NDI</b>	Pearson Correlation coefficient	1	0.544	0.347
	p-value		< 0.001	0.001
<b>Decrease in HADS depression score</b>	Pearson Correlation coefficient	0.544	1	0.552
	p-value	< 0.001		< 0.001
<b>Decrease in HADS anxiety score</b>	Pearson Correlation coefficient	0.347	0.552	1
	p-value	0.001	< 0.001	

\* Correlation is significant at the 0.01 level (2-tailed).

\*\* HADS = Hospital Anxiety and Depression Scale, NDI = Neck Disability Index



**Figure 7.** Marginal NDI over time (in blue) with a 95 % confidence interval (dark gray). The NDI of individual subjects are plotted in the background.



**Figure 8.** NDI during follow-up for the Delta NDI groups.

*Patients were grouped based on their HADS score over time.*

## 2.2 Explanatory Linear Mixed Effects Model

To quantify the effect of HADS on NDI over time, follow-up moment and deviation from the mean HADS were used in a LMM (Table 6). The intercept of 41.16 can be interpreted as the NDI at baseline for a patient with an average HADS score, with a decrease of 20.68 points in NDI after one year compared to baseline or 19.92 after two years compared to baseline. All predictors are significant ( $p < 0.001$ ) (Table 6). Deviation HADS has a beta coefficient of 1.34, meaning that one point increase of HADS compared with the mean HADS results in an NDI increase of 1.34, regardless of time. The ICC is 0.24, lower than for the predictive LMM, most like due to the fact



that the NDI at baseline is an outcome in this model and not a predictor, highlighting the previously discussed importance of baseline NDI for accurate estimation of NDI over time.

**Table 6.** Coefficients of explanatory Linear Mixed Effects Model with its corresponding standard error (SE), degrees of freedom (DF), t-value and p-value. HADS is centered at its mean.

	<b>Coefficients</b>	<b>SE</b>	<b>DF</b>	<b>t-value</b>	<b>p-value</b>
(Intercept)	41.16	1.32	187.00	31.25	<0.001
Time: 52 weeks	-20.68	1.45	187.00	-14.31	<0.001
Time: 104 weeks	-19.92	1.45	187.00	-13.75	<0.001
Deviation HADS	1.34	1.34	187.00	11.67	<0.001

## Discussion

Patients that were classified as depression or anxiety cases at baseline had statistically significant and clinically relevant higher NDI scores one and two years after surgery. The crude NDI averages for the delta HADS groups illustrate that also during follow-up NDI changes in the same direction as the HADS group changes. Additionally, a predictive LMM and explanatory LMM enabled the successful analysis of decreased mental health using HADS on a continuous scale, as it relates to functional outcome in NDI respectively over time. Lastly, a method was proposed to effectively communicate results from the predictive LMM, using an R-shiny application.

These results raise the question whether decreased mental health status is either a patient characteristic or comes inherently with experiencing pain and disability from cervical disc disease. Whether disability causes symptoms of decreased mental health or whether decreased mental health causes patients to experience higher disability, is not assessed in this study. A combination of both is assumed to be most likely, which would support the recommendation for future research to assess if preoperative treatment of depressive or anxious symptoms could improve functional outcome after surgery, both immediately postoperative as well as long-term. This has only been shown in one previous study on a relatively small number of patients that used a pharmaceutical intervention [26]. Other than pharmaceutical intervention, preoperative counselling or cognitive behavioural therapy could be other potentially interesting strategies to investigate.

The use of prospectively collected, high quality data from an RCT is a major strength of this study. This allowed for repeated measurement analysis of the HADS scores and assessment of their predictive value. Furthermore, it provided the opportunity to analyse HADS scores on a continuous scale, rather than dichotomizing results and losing the granular aspect of this outcome scale, as has been done previously [14]. Another strength is the direct implementation of predictive modelling results into an application. Improving presurgical counselling possibilities in this manner is increasingly important with the rise of shared decision making in the current medical world, for which effective communication of research results is paramount.

However, this study has limitations. A limitation for the external validity of this study is the exclusion of patients with severe mental and psychiatric disorders from the NECK-trial. However, it could be argued that excluding severely depressed patients strengthens our conclusions, as the effect was illustrated in patients suffering from relatively 'mild' symptoms.

The use of only clinical, without radiological or histological, parameters is another limitation to this study and was illustrated by the percentage of within-group variance that could not be explained. The explanatory LMM showed that using only HADS scores explained 24% (ICC 0.24) of the within-group variance of NDI scores, whereas the predictive LMM, that incorporated baseline NDI score as well, adds another 51% (ICC 0.75). However, there remains to be 25% of variance unexplained and therefore we plea to combine different types of outcome parameters, as well as increasing the sample size of analysed patients, in future research in order to achieve higher accuracy in predictions and bring the percentage of unexplained variance down. Radiological imaging data has recently been successfully used in cervical spine disease to predict outcomes with Deep Learning techniques [27]. Other studies have shown how in sciatica the histopathological parameters in disc tissue, such as different types of pro-inflammatory cytokines, could be associated with worsened pain symptoms [28]. However, the combination of both radiological, clinical and histological parameters could draw an even more complete picture of the patient and can therefore be expected to achieve higher accuracy in predicting outcomes for individual patients.

Moreover, the duration of follow-up is another limitation, as the analysed data was collected at baseline, one and two years after surgery, but ideally clinical predictions would be made for long-term outcomes, five to ten years after surgery.

Lastly, in future causality research, determining the direction of the effect between mental health and disability scores after cervical spine surgery should be addressed, as it may provide additional insights in how to manage patients with mental illnesses before and after spine surgery.

Patients suffering from depression and anxiety before cervical spine surgery demonstrate significantly more neck disability one and two years after surgery, and therefore don't benefit from surgery in the same way other patients do. Additionally, this study demonstrates that if, during follow-up, symptoms of depression and anxiety improve, patients their functional outcome improves as well. Using predictive modelling, it was additionally shown that mental health can be used to explain and predict the changes in neck disability after surgery. Lastly, an R-shiny application was developed to facilitate an easier-to-interpret, visual communication of these models to patients during a preoperative clinic visit.

Using applications, like the one designed in this study based on the predictive modelling developed, can aid personalized treatment counselling and is a promising development for future shared decision making healthcare.

## References

1. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and Pain Comorbidity: A Literature Review. *Archives of Internal Medicine*. 2003;163(20):2433-45.
2. Lamb SE, Guralnik JM, Buchner DM, Ferrucci LM, Hochberg MC, Simonsick EM, et al. Factors that modify the association between knee pain and mobility limitation in older women: the Women's Health and Aging Study. *Ann Rheum Dis*. 2000;59(5):331-7.
3. Wells KB, Golding JM, Burnam MA. Affective, substance use, and anxiety disorders in persons with arthritis, diabetes, heart disease, high blood pressure, or chronic lung conditions. *General Hospital Psychiatry*. 1989;11(5):320-7.
4. Betrus PA, Elmore SK, Hamilton PA. Women and somatization: Unrecognized depression. *Health Care for Women International*. 1995;16(4):287-97.
5. Burton AK, Tillotson KM, Main CJ, Hollis S. Psychosocial Predictors of Outcome in Acute and Subchronic Low Back Trouble. *Spine*. 1995;20(6).
6. Gureje O, Simon GE, Von Korff M. A cross-national study of the course of persistent pain in primary care. *PAIN*. 2001;92(1).
7. Potter RG, Jones JM. The evolution of chronic pain among patients with musculoskeletal problems: a pilot study in primary care. *Br J Gen Pract*. 1992;42(364):462-4.
8. Reis S, Hermoni D, Borkan JM, Biderman A, Tabenkin C, Porat A. A new look at low back complaints in primary care: a RAMBAM Israeli Family Practice Research Network study. *J Fam Pract*. 1999;48(4):299-303.
9. Jackson KL, Rumley J, Griffith M, Agochukwu U, DeVine J. Correlating Psychological Comorbidities and Outcomes After Spine Surgery. *Global Spine Journal*. 2019;10(7):929-39.
10. Edwards RR, Klick B, Buenaver L, Max MB, Haythornthwaite JA, Keller RB, et al. Symptoms of distress as prospective predictors of pain-related sciatica treatment outcomes. *Pain*. 2007;130(1-2):47-55.
11. Adogwa O, Elsamadicy AA, Cheng J, Bagley C. Pretreatment of Anxiety Before Cervical Spine Surgery Improves Clinical Outcomes: A Prospective, Single-Institution Experience. *World Neurosurg*. 2016;88:625-30.
12. Elsamadicy AA, Adogwa O, Cheng J, Bagley C. Pretreatment of Depression Before Cervical Spine Surgery Improves Patients' Perception of Postoperative Health Status: A Retrospective, Single Institutional Experience. *World Neurosurg*. 2016;87:214-9.
13. Alvin MD, Miller JA, Sundar S, Lockwood M, Lubelski D, Nowacki AS, et al. The impact of preoperative depression on quality of life outcomes after posterior cervical fusion. *Spine J*. 2015;15(1):79-85.
14. Toru D, Hideki N, Koji N, Shima H, Yusuke S, So K, et al. Effect of depression and anxiety on health-related quality of life outcomes and patient satisfaction after surgery for cervical compressive myelopathy. *Journal of Neurosurgery: Spine SPI*. 2019;31(6):816-23.
15. Arts MP, Brand R, van den Akker E, Koes BW, Peul WC. The NETHERLANDS Cervical Kinematics (NECK) Trial. Cost-effectiveness of anterior cervical discectomy with or without interbody fusion and arthroplasty in the treatment of cervical disc herniation; a double-blind randomised multicenter study. *BMC Musculoskelet Disord*. 2010.
16. Vleggeert-Lankamp CLA, Janssen TMH, van Zwet E, Goedmakers CMW, Bosscher L, Peul W, et al. The NECK trial: Effectiveness of anterior cervical discectomy with or without interbody fusion and arthroplasty in the treatment of cervical disc herniation; a double-blinded randomized controlled trial. *Spine J*. 2019;19(6):965-75.

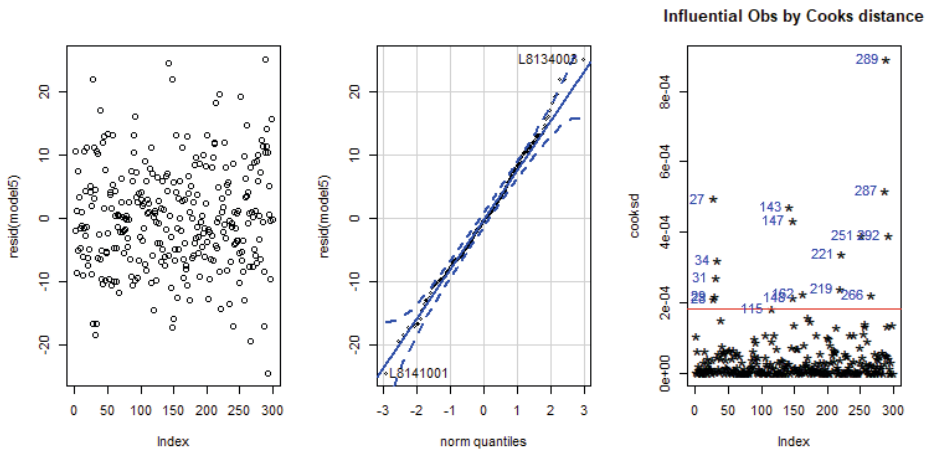
17. Herrmann C. International experiences with the Hospital Anxiety and Depression Scale--a review of validation data and clinical results. *J Psychosom Res.* 1997;42(1):17-41.
18. Mykletun A, Stordal E, Dahl AA. Hospital Anxiety and Depression (HAD) scale: factor structure, item analyses and internal consistency in a large population. *Br J Psychiatry.* 2001;179:540-4.
19. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther.* 1991;14(7):409-15.
20. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF. Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. *Spine (Phila Pa 1976).* 2002;27(5):515-22.
21. Vos CJ, Verhagen AP, Koes BW. Reliability and responsiveness of the Dutch version of the Neck Disability Index in patients with acute neck pain in general practice. *Eur Spine J.* 2006;15(11):1729-36.
22. Auffinger BM, Lall RR, Dahdaleh NS, Wong AP, Lam SK, Koski T, et al. Measuring surgical outcomes in cervical spondylotic myelopathy patients undergoing anterior cervical discectomy and fusion: assessment of minimum clinically important difference. *PLoS One.* 2013;8(6):e67408.
23. Lauche R, Langhorst J, Dobos GJ, Cramer H. Clinically meaningful differences in pain, disability and quality of life for chronic nonspecific neck pain - a reanalysis of 4 randomized controlled trials of cupping therapy. *Complement Ther Med.* 2013;21(4):342-7.
24. Parker SL, Godil SS, Shau DN, Mendenhall SK, McGirt MJ. Assessment of the minimum clinically important difference in pain, disability, and quality of life after anterior cervical discectomy and fusion: clinical article. *J Neurosurg Spine.* 2013;18(2):154-60.
25. Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. *J Orthop Surg Res.* 2015;10:24.
26. Bender R. Calculating confidence intervals for the number needed to treat. *Control Clin Trials.* 2001;22(2):102-10.
27. Goedmakers CMW, Lak AM, Duey AH, Senko AW, Arnaout O, Groff MW, et al. Deep Learning for Adjacent Segment Disease at Preoperative MRI for Cervical Radiculopathy. *Radiology.* 2021;301(3):664-71.
28. Djuric N, Lafeber GCM, Vleggeert-Lankamp CLA. The contradictory effect of macrophage-related cytokine expression in lumbar disc herniations: a systematic review. *European Spine Journal.* 2020;29(7):1649-59.

## Appendix

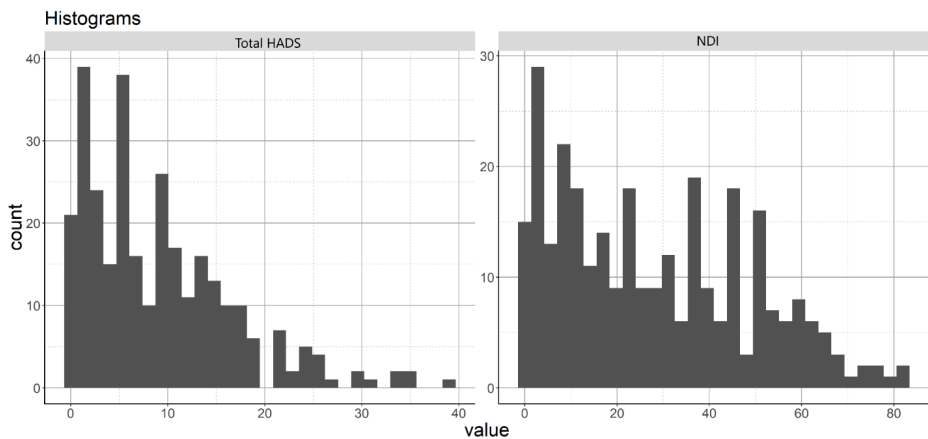
### A Code

All code can be found on the Github [https://github.com/ilsevb95/Stat\\_Consulting](https://github.com/ilsevb95/Stat_Consulting)

### B Model Assumption plots for explanatory LMM, from left to right: residuals plot (homogeneity of variance), QQ-plot (normality), Cook's distance (no outliers)

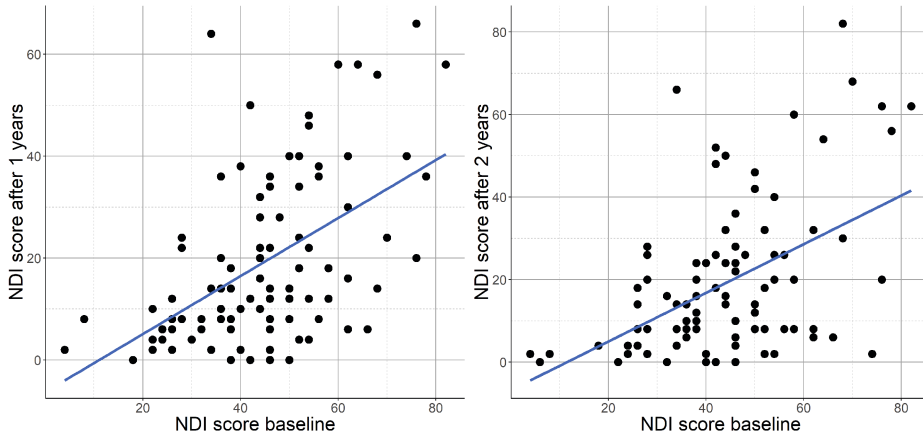


### C Histograms of distribution for NDI & HADS total scores



5

**D Correlation plots of NDI at baseline with NDI after 1 and 2 years**



**E Crude average NDI with standard deviation per delta HADS group**

Delta HADS group	Baseline	52 Weeks	104 Weeks
No (doubtful) case at baseline, (doubtful) case after two years	39.14 ± 12.32	22.86 ± 24.46	27.71 ± 23.11
No (doubtful) case at baseline, no (doubtful) case after two years	40.14 ± 14.07	14.30 ± 12.65	13.22 ± 11.63
(Doubtful) case at baseline, no (doubtful) case after two years	47.55 ± 13.56	20.45 ± 16.88	13.40 ± 11.77
(Doubtful) case at baseline, (doubtful) case after two years	58.80 ± 16.19	34.14 ± 19.09	44.13 ± 22.29