SAHLGRENSKA ACADEMY

Automatic analysis of intervertebral discs based on deep learning – comparing preoperative MR images with 1-year post lumbar disc herniation surgery outcomes.

Degree Project in Medicine

Emil Cedergårdh

Programme in Medicine

Gothenburg, Sweden 2020

Supervisors:
Helena Brisby, Professor
Kerstin Lagerstrand, Assoc. Professor
Olof Westin, MD, PhD.

Department of Orthopaedics at Institute of Clinical Science
Sahlgrenska Academy and Sahlgrenska University Hospital

Table of Contents

Abstract

Degree project. Program in Medicine. ***Automatic analysis of intervertebral discs based on deep learning – comparing preoperative MR images with 1-year post lumbar disc herniation surgery outcomes*****. Emil Cedergårdh, 2019. Institute of Clinical Science, Department of Orthopedics. Gothenburg, Sweden.**

Introduction: Lumbar disc herniation surgery often leads to major improvement in leg pain; however, some patients have remaining back pain that might depend on the level of disc degeneration. Magnetic resonance imaging (MRI) examination has a central role in the preoperative evaluation. Today, the images are reviewed by a radiologist, a task which, in the future, might be assisted by computers using artificial intelligence (AI) and deep learning. In this study, intervertebral disc (IVD) characteristics from preoperative MR images is extracted and then compared with the 1-year post lumbar disc herniation surgery outcome. Due to technical error, semi-automated segmentation was used instead of deep learning-based segmentation.

Aim: To study if there is a relationship between midsagittal signal intensity measures in preoperative MR images and 1-years postoperative patient reported outcome measures (PROM´s) on back pain, physical function and overall satisfaction.

Method: Patients undergoing lumbar disc herniation surgery at Sahlgrenska University Hospital during the years 2013-2017 (n=218) and registered in the Swedish National Quality Registry for Spine Surgery (Swespine) were included. In each patient, the midsagittal part of the herniated IVD was segmented (outlined) on preoperative T2-weighted MR images using an in-house developed software. Signal intensity measures were calculated and statistically compared (t-test at $p<0.05$) to the PROM´s Numeric rating scale (NRS) back, Oswestry disability index (ODI) and Global Assessment (GA).

Results: No significant difference in signal intensity measures between patients with successful versus unsuccessful PROMS´s was found.

Conclusions: This study could not prove any relationship between midsagittal signal intensity measures in preoperative MR images and 1-years postoperative PROM´s. Further studies are encouraged using standardized MRI protocol and scanner, and more patient's data enabling adjustment of confounders.

Key words: Lumbar Disc herniation, Automatic Segmentation, Deep Learning, MRI, Signal Intensity

Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Networks
GA	Global Assessment
IVD	Intervertebral Disc
MRI	Magnetic Resonance Imaging
MSI	Mean Signal Intensity
NRS	Numeric Pain Rating Scale
ODI	Oswestry Disability Index
PROM´s	Patient Reported Outcome Measures
ROI	Region of Interest
SDSI	Standard Deviation of Signal Intensity
Swespine	The Swedish National Quality Registry for Spine Surgery

Introduction

Lumbar disc herniation is a common condition which has considerable impact on individual's everyday life and healthcare resources. The prevalence of symptomatic lumbar disc herniation is reported to be 1-3% depending on age, country and gender (1). Associated symptoms are radiating leg pain along the sciatic nerve sometimes in connection with lower back pain. Lumbar disc herniation is either conservative or surgical treated and, just in Sweden, more than n=2000 patients undergoes lumbar disc herniation surgery annually (2). Disc herniation is a clinical diagnose set by physical examination with complementary Magnetic Resonance Imaging (MRI) examination (3, 4). At present, the MR images are reviewed by a radiologist and their opinion is used in the preoperative evaluation. If the clinical findings correlate to the MRI findings and the leg pain do not subside within the first months surgical treatment may be in question (5).

In general, surgical treatment of disc herniation leads to fast pain-relief and a majority (75%) of the patients are satisfied with the surgical outcome (2). However, some patients experience less relief in back pain, compared to leg pain, and the reoperation rate is reported being 15%, eight years after surgery (6). Why some patient experience limited relief in back pain or need reoperations is questionable and might depend on the grade of disc degeneration before surgery, thus, possible to detect with MRI.

At present, the utilization of the MRI technique is limited to the human eye and the knowledge and experience of the radiologist. Thanks to current increase in available compute power, the analysis process can be automatically executed with artificial intelligence and deep

learning (7), i.e. with computer system able to solve task normally requiring human intelligence. The technique might enable robust and accurate MRI diagnostics and, thus, have potential to transform the preoperative evaluation of patients with disc herniation (8). Earlier studies have shown good result in automatic detection of radiologic features using deep learning (9). However, to our knowledge, no one has compared automatic produced features with clinical postoperative outcomes.

In this retrospective study, patients from Swedish national quality registry for spine surgery (Swespine) is studied. Intervertebral disc (IVD) characteristics, volume and signal intensity, from preoperative MR images will be produced by a software for automatic segmentation, based on deep learning. These characteristics will be compared to the patient reported outcome measures (PROM´s) 1-year post lumbar disc herniation surgery.

The aim was to study if there is a relationship between the IVD characteristics and the postoperative PROM´s. If a relationship indeed exists and the IVD characteristics can be used for prediction of surgical outcome, IVD analysis based on automatic segmentation of the MR images might be a useful tool in future preoperative evaluation by increasing the throughput and the utilization of the MRI examination.

The Human Spine – Anatomy

The human spine, or vertebral column, consists of 7 cervical, 12 thoracic and 5 lumbar vertebrae. Each vertebra has a vertebral body and a vertebral arch. Posterior of the vertebra body and anterior of the arch is an opening, called the vertebral foramen. The succession of vertebral foramina forms the vertebral canal holding the spinal cord, which is a continuation of the medulla oblongata and connects the brain with the body. The spinal cord is made of 31 segments from which symmetrical pair of nerve branches, holding both motor and sensory nerves, forming the spinal nerves. The spinal nerves leave the vertebral canal through the intervertebral foramen at each vertebral level. All vertebrae are separated by a disc shaped structure named intervertebral disc (IVD) which enables movement, withstand and transfer loads of the spine. (10)

The movement of the spine is limited by different ligaments and the shape and disposition of the facet joints (the joint between each vertebrae). The vertebrae of different segments (cervical, thoracic and lumbar) of the human spine have different characteristics. The lumbar vertebrae, named L1-L5, are characterized by their greater size giving them ability to stand the heavy load of the upper body. The facet joints of lumbar spine are oriented in a way allowing flexion, extension and lateral flexion but prohibits rotation. (10)

The intervertebral disc - IVD

The IVD is a disc-shaped structure measuring 7-10 mm thickness and 40 mm in diameter in the lumbar region and consist of an outer fibrosus part, called annulus fibrosus, and a gelatinous central mass, called nucleus pulposus (11). The difference between its components

is mainly the proportions of collagen. High content of type 1 collagen makes annulus fibrosus a strong fibrous ring able to withstand heavy loads (12). The nucleus pulposus contains up to 90% water due to a matrix of type 2 collagen and water binding proteoglycan molecules, which makes it a viscoelastic structure. During different movements of the spine, e.g. flexion, the nucleus pulposus move posterior (during flexion) towards the annulus fibrosus. Between the IVD and the vertebral body, both superiorly and inferiorly, is a less than 1 mm thick horizontal layer of hyaline cartilage, called the endplate. The endplates absorb the hydrostatic pressure created by axial load of the spine and prevent the nucleus from bulging into the successive vertebrae. Diffusion of solutes across the endplates serve the IVDs with nutrients (13).

Already early in life, the IVDs start to degenerate. Most probably, this is due to a physiological decrease in blood supply through the endplate, leading to tissue breakdown in the nucleus pulposus (14) or by physical disruption in the annulus (15). Loss of proteoglycan is the most significant biochemical change in the IVD associated with degeneration. The process makes the IVD less hydrated due to a fall in osmotic pressure. Degeneration is also associated with a loss of collagen fiber and a decrease in number of viable cells in the nucleus (11, 16). However, this is less obvious in MRI. Additionally, a sign of disc degeneration is the formation of tears in the anulus fibrosus (17) that may lead to low back pain (18).

Disc Herniation

Spinal disc herniation is by definition when parts of the disc, often nucleus pulposus, displace outside its normal limits and is a common injury often caused by a disruption in anulus fibrosus (19) or, possibly more commonly, in the endplate (20). Radiological terminology distinguishes among disc bulging, protrusion, extrusion and sequestration where disc bulging is not considered a form of herniation (21). A herniation, still in connection with the disc, is classified as an extrusion if the greatest measure of the displaced disc material is greater than the measure of its base, else a protrusion. If the displaced disc material is completely separated from the disc, the herniation is classified as a sequestration (19).

In a herniation located at the posterolateral aspect of the disc, protruding nucleus pulposus may cause mechanical pressure on the transverse nerve root inferiorly of the disc, which results in pain. An inflammatory process probably caused by leakage of the nucleus pulposus also contribute to the pain (22). A majority, at least 95%, of all herniated discs are located at level L4-L5 or L5-S1 (23, 24). High compression load of the spine with simultaneous flexion have in experimental models shown to cause disc herniations (25).

The primary signs and symptoms of lumbar disc herniation are radicular pain along the sciatic-nerve distribution, in the buttocks, thigh and calf, called “Sciatica” (22). Acute or more slowly progressive lower back pain can also be seen. Sensory and or motoric loss, corresponding to the affected nerve root, is also a characteristic. The distribution of the pain and the functional loss is dependent on the level and location of the herniation. A paracentral herniation at L4-L5 would affect the traversing nerve root causing L5 radiculopathy whereas a

lateral herniation, at the same level, would affect the exiting nerve root causing L4 radiculopathy (22). Because of higher disc pressure, patients often report increased pain while sitting (26). In rare cases with massive herniation and compression of cauda equina, symptoms as saddle anesthesia, urinary incontinence or retention and loss of anal sphincter tonus may be seen (21). Cauda equina syndrome is an acute indication for further radiological examination and treatment.

The clinical diagnosis of lumbar disc herniation with radiculopathy is set by the patient ´s history in combination with manual muscle testing, sensory testing, and straight leg rise test (SLR) or Lasegues test. A meta-analysis from 2017 concluded a positive SLR test together with 3 out of 4 of following symptoms: dermatomal pain along a nerve root, sensory deficit, reflex deficit and/or motor weakness meets the requirements for diagnose (3).

Treatment

Patients with sciatica and a suspected lumbar disc herniation, should primarily undergo conservative therapy, often including anti-inflammatory drugs and exercise-based physical therapy. Studies have shown a majority of disc herniations resolve naturally without surgery (27). MRI, which is the best radiological examination for detecting disc degeneration (4), is indicated if the patient experiences no response or pain relief within 6 weeks (4, 28). However, since disc degeneration including disc bulging and disc protrusion is also common among asymptomatic persons, radiological findings must be correlated to clinical signs (5, 29, 30). If so, the patient might be candidate for surgery. Surgery has shown major benefits

compared to conservative treatments when it comes to fast relief of sciatica. Regarding back pain relief, however, only a smaller advantage can be seen (24, 31).

Surgery for disc herniation was first performed in the early 1900s (32) and the popularity of disc surgery increased rapidly in 1934 when the correct pathogenesis, and an appropriate surgical treatment, of disc herniation was described (33, 34). Today, there are many different types of lumbar disc herniation surgery techniques, all with the basic principle to relieve nerve root compression with removal of the herniated part of the disc (35). Microdiscectomy has long been the most common procedure for lumbar disc herniation and is, in general, a normal open discectomy, guided by microscope enabling a smaller incision with less dissection (36). However, modern magnification and illumination systems with microscope, and in the last decade endoscope, have actualized minimal invasive techniques that reduce incision size and area of dissection with less soft tissue injury (37). Today, there are several known percutaneous endoscopic approaches, including interlaminar and transforaminal (22), but still, the gold standard of surgical management for lumbar disc herniation is open discectomy with partial laminectomy (38). Multiple studies have compared outcome of different minimal invasive techniques without or with limited conclusions which technique is better than others (39-42).

In general, 1-year follow-ups after disc herniation surgery show good results with great decrease in back, as well as leg, pain. The majority (75%) of the patients are satisfied with the surgical outcome and the overall disability from lower back pain is significantly decreased. However, some of the patients are uncertain (18%) or dissatisfied (7%) with the surgical outcome and approximately 10% of the patients rate their back pain as well as leg pain unchanged or worsened 1 year after surgery. This group of uncertain or dissatisfied patients

might have got a non-optimal treatment and highlights the importance of good preoperative evaluation with high specificity and sensitivity. (2)

Magnetic resonance imaging – MRI

Magnetic resonance imaging (MRI) is a medical imaging technique used in radiology to characterize tissue changes, including disc degenerations (21). The technique is based on the magnetic resonance phenomenon. That is, the hydrogen nuclei in the body that is magnetized by the strong magnetic fields of the MRI scanner will be affected (excited) by radio frequency fields at resonance. The signal of the excited hydrogen nuclei will differ between hydrogen nuclei in water molecules and lipids and between tissues of different structures. This is due to the fact that excited nuclei in different tissues and hydrogen compounds display different time factors for their return to the ground state, so called T1 and T2 relaxation times (values). Water and cerebrospinal fluid (CSF) have long T1 values (3000-5000 ms) and appear dark in T1-weighted images, while fat appears bright because of short T1 time (260 ms) (43). In T2-weighted images, both water and fat appear bright. Most pathological processes show increased T1 and T2 times. Hence, they become dark in T1-weighted and bright in T2-weighted images. (44)

Increased T2-weighted signal in the posterior part of the IVD highly suggest disc herniation (45). The MRI findings of disc degeneration is often classified by a radiologist using the Pfirrmann grade (46), a five-step grading scale manually divided by different characteristics

of the IVD such as homogeneity, signal intensity and height. The mean T2-weighted signal intensity of the IVD (47) and the standard deviation of the mean signal (48) enable continuous quantitative measures of disc degeneration and the signal intensity correlates with the level of proteoglycan content in the IVD (49). These measures were analyzed in this study.

Artificial Intelligence – Deep Learning

Artificial intelligence (AI) is the theory and development of a computer system able to solve tasks normally requiring human intelligence. Major progress is being made within the field thanks to an explosion of the available compute power (7) and today, AI is used in many different systems among visual perceptions, speech recognition, translating between languages and many more. Machine learning is a subset of AI and the scientific study of algorithms computer systems use to perform task without preprogrammed instructions with ability to process large and complex datasets were statistical analysis would be unfeasible. Machine learning can be divided into supervised and unsupervised learning. Supervised learning means learning from labeled information in a training set of data where the inferred function can be used for mapping new data. Unsupervised learning tries to find hidden patterns present in datasets without the need of labeled information, more particularly without manual guidance (7, 50). Furthermore, as a subset of machine learning, deep learning use artificial neural networks to mimic the synaptic connections in the human brain with multiple layers of information processing enabling task solving by learning from experience (51). The technology is still in its infancy but may be used to predict cardiovascular health from fundus

images of the retina (52) or identifying melanoma from images of skin lesions (53). Automatic detection segmentation of abnormalities from lumbar MRI has been described as a difficult task due to partial volume effect, where multiple tissues contributes to pixels and blurs, and intensity inhomogeneities, where the same tissue gives rise to different intensity variation (54).

In medical imaging the use of deep learning is mostly by a kind of artificial neural network called convolutional neural networks (CNN). The CNN is designed to arrange the image (the input) in a grid structure and then feed it through multiple layers of convolutions and activations with few connections between the layers (7). While processing the grids the CNN preserve the spatial relationship in the data. By now, CNN have surpassed even human performance in visual image recognition (55) and in detection of radiological features, CNN shows comparable results with an expert radiologist (9).

Swespine

Swespine, i.e. the Swedish national quality register for spine surgery, was launched in 1993 and is currently holding more than 125,000 index operations. The register covers 98% of the clinical departments in Sweden and data on approximately 75%-80% of all patients undergoing spinal surgery are reported into the register. During the last years, approximately 10,000 operations have annually been registered in Swespine. The information is collected by forms completed by the surgeon, at the operation, and by the patients who complete

questionnaires both pre- and postoperatively (at 1-, 2-, 5- and 10-years). The size of the register and the good coverage makes Swespine an valuable source for research; only in 2017 12 articles based on Swespine data were published. (2)

Aim

In this study, IVD characteristics in preoperative MR images are analyzed. The aim was to investigate if any relationship between these characteristics and PROM´s reported in Swespine register 1-year post lumbar disc herniation surgery could be detected.

Research question

Is there a relationship between IVD characteristics on preoperative MR images, such as signal intensity and variance measures - extracted from automatic segmentations based on deep learning, and 1-years postoperative PROM´s; Numeric pain rating scale (NRS) back, Global assessment (GA) and Oswestry disability index (ODI)?

Is there a relationship between midsagittal standard signal intensity in herniated IVDs on preoperative MR images and 1-years postoperative PROM´s; NRS back, GA and ODI?

Material and Methods

Study population

In this retrospective study, a total of n=375 patients were retrieved from Swespine. The patients have all gone through conventional, microscopic or endoscopic lumbar disc herniation surgery at Sahlgrenska University Hospital, Gothenburg, during the period from January 2013 to December 2017. Patient with no 1-year follow up records were excluded (n=143). By matching the patients age and date for surgery, retrieved from Swespine with the in-house surgery schedule program Operätt, full social security number of the patients was found for all but 11 patients and used to localize the preoperative MRI investigations. The latest preoperative spinal MRI from each of the patients (n=221) was extracted from the database of the Region Västra Götaland. In n=3 cases, no spinal MRI series was found, and the patients were excluded. To validate the automatic segmentation method, n=13 patients were randomly selected for manual segmentation and thereby excluded from the study population. Of the 205 patients, the group consisted of n=109 (53.2%) women and n=96 (46.8%) men, all with mean age of 42.6 years at the time of surgery. *Figure 1* shows the included and excluded patients of the study.

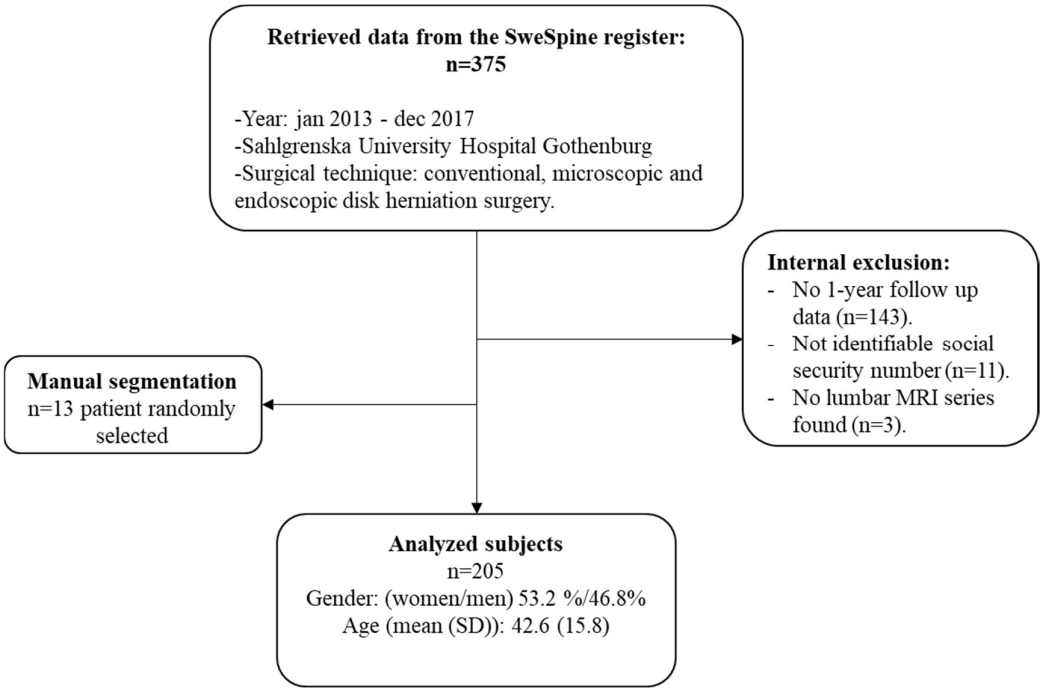

Figure 1: Flow-chart of included and excluded patients.

Surgery outcomes

Multiple validated outcome measures are used to monitor status of health and treatment effect after lumbar spine surgery. In this study, three different 1-year outcome values were used, all described in the paragraphs below. Back pain was measured using the numeric pain rating scale (NRS). Treatment effect and patient functional status were evaluated with global assessment (GA) of back pain and Oswestry Disability Index (ODI). In our study population all patients have reported NRS-back at the 1-year follow-up. However, records of GA was missing in n=1 patient and ODI was missing in n=6 patients.

NRS is an 11-point scale were the patients might rate their pain where 0 means “no pain” and 10 means “worst imaginable pain”. NRS is a validated method to measure pain (56) and is the most common pain outcome measure in chronic low back pain patients (57). Studies have examined the use of NRS on low back pain patients and concluded a 2-point change on the NRS is a clinically meaningful change beyond statistically measurement errors (58).

Global Assessment (GA) is used as a basic reference in studies of responsiveness to a change and is based on a single question about treatment effect, “How is you back pain today as compared to before the surgery?”. The patients answer with six options generating a score from 0 to 5 (59). *Table 1* shows the GA questionnaire. Studies have shown that GA is a valid and responsive descriptor of the overall effect of lower back pain treatment (59, 60).

Table 1: Global Assessment questionnaire.

How is your backpain today as compared to before the surgery?
0 = I had no pain prior to surgery
1 = Completely pain-free
2 = Significantly improved
3 = Somewhat improved
4 = Unchanged
5 = Deteriorated

The Oswestry Disability Index (ODI) is a patient-completed questionnaire used to quantify disability of low back pain. The questionnaire is based on questions about ten topics; pain intensity during movements, personal care, lifting capacity and the ability of walking, sitting, standing, sleeping, sex life, social life and travelling. Each topic gives a maximum of 5 point.

A patient's total score is calculated and ODI is the total score in percentage of the maximum 50 point, where 0-20% indicates minimal disability and 81-100% indicates a bedbound patient. The ODI is the most frequently used functional outcome measures in chronic low back pain patients (57). *Appendices* 1 show the ODI questionnaire. (61)

Automatic segmentation and IVD characteristics

Segmentation of an image means dividing it into different regions of interest (ROI), in this case IVDs (8). To validate the segmentation method, n=13 out of n=221 T2-weighted MR series were randomly selected, but with the restriction of including all clinical MRI scanners used in the examinations of the cohort. The manual segmentation was performed using the software ITK-SNAP version 3.8.0 (62). ITK-SNAP works as a painting tool where it is possible to color any region of interest in MR images, in this case lumbar IVD, pixel by pixel (*Figure 2*). In each selected MR series, representing a patient, all n=5 lumbar IVD was segmented, thus colored and labeled. The manually segmented MRI series were then used as a training and validation set for the segmentation method, using a software of convolutional neural networks, based on deep learning.

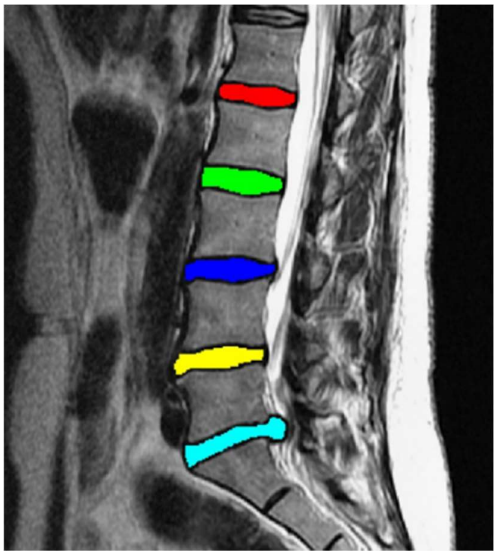

Figure 2: Manual segmentation of 5 consecutive intervertebral discs on T2-weighted magnetic resonance images, using the software ITK-SNAP.

The original idea was to let the software identify all lumbar IVDs of the patients in the study population and automatically perform the segmentation of the IVDs. Then, three IVD characteristics were supposed to be generated by the software; disc volume, mean signal intensity and standard deviation of the mean signal intensity. The two latter as a measure of disc degeneration (63). Unfortunately, the software did not perform as good as desired for herniated IVDs and no data of IVD characteristics could be generated with this software tool. As a result, the research question was rephrased and the methodology regarding comparing IVD characteristics with surgical outcomes was changed (Material and Methods – Part 2).

Material and Methods – Part 2

Study population – Part 2

In Part 2, all patients with available preoperative MRI were included, thus no patient were excluded for manual segmentation, as earlier. The study population of n=218 patients consisted of n=113 (51.8%) women and n=105 (48.2%) men, all with mean age of 42.4 years at the time of surgery. *Figure 3* shows included and excluded patients.

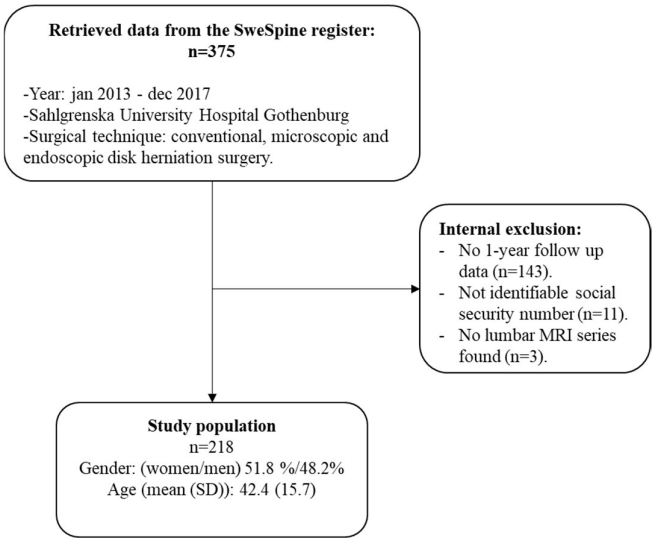

Figure 3: Flow-chart of included and excluded patient in Part 2.

Surgery outcomes – Part 2

The same PROM´s (NRS, GA and ODI) as earlier were used in Part 2 (see section *Surgery outcomes)*.

In the study population of n=218 patients NRS back were reported in all patients. However, records of GA was missing in n=1 patient and ODI was missing in n=6 patients. Dichotomization was made for each outcome based on successful surgery outcomes (Group 0), thus patients with limited symptoms or disability, versus unsuccessful outcome (Group 1), thus patients with symptoms. NRS back were dichotomized in ≤2 versus >2, GA in ≤2 versus >2 and ODI in ≤20 versus >20. This generated three different subgroupings of the study population, each handled separately in the statistical analysis.

Segmentation and IVD characteristics – Part 2

For all patients, each MRI series, was post processed using an inhouse-developed software based on MATLAB (R206a, Mathworks®, Natick Massachusetts, USA), as in previous Gothenburg based study (64). Each herniated IVD was outlined on three consecutive T2-weighted midsagittal slices using semi-automated segmentation (*Figure 4*). In n=8 patients with records of two herniated IVDs, the most symptomatic IVD, according to the medical records, was analyzed. In order to evaluate regional characteristics, each IVD was divided into 5 equally sized subregions (ROI), based on the total midsagittal length of the IVD, ranging from 1 (anterior) to 5 (posterior) (*Figure 5*). Same method has been used in earlier studies (65, 66). From the manual segmentation, mean signal intensity (MSI), standard

deviation of signal intensity (SDSI) and SDSI/MSI of the midsagittal part of the IVD were calculated.

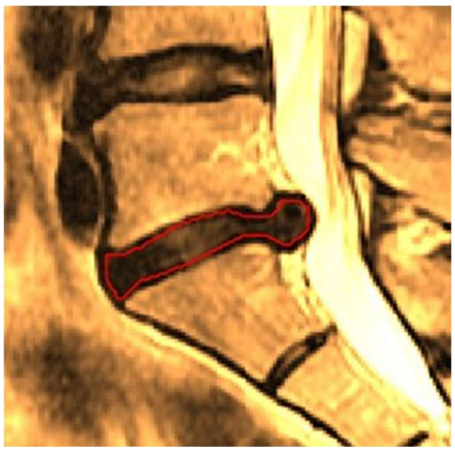

Figure 4: Semi-automated segmentation of disc with a herniation (L5-S1) on T2-weighted magnetic resonance images, using MATLAB software R206a

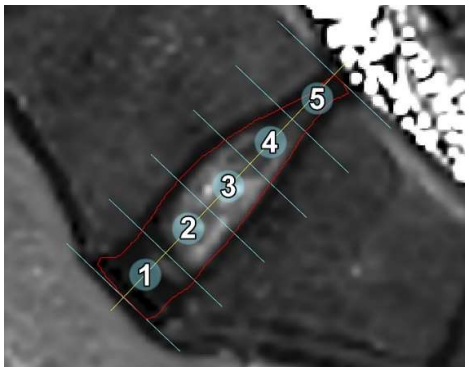

Figure 5: Example of the segmentation performed on a L5-S1 intervertebral disc magnetic resonance image *section. Each segmentation was divided into five equally sized subregions, ranging from 1 (anterior) to 5 (posterior), using MATLAB software R206a (Mathworks®, Natick Massachusetts, USA). Picture from Waldenberg et. al. (65).*

Data analysis and statistical methods – Part 2

The statistical software IBM SPSS version 25 was used for statistical analysis. Independent samples t-test were used to evaluate if there was a significant difference in the different signal intensity measures between patients with successful versus unsuccessful surgical outcome. Measures of the whole midsagittal part and of the different subregions were compared separately.

Ethics

This study got ethical approval from the Regional Ethical Review Board Gothenburg at Sahlgrenska Academy, University of Gothenburg, Sweden (DNR 753-17). All patients got both oral and written information about data collections in Swespine and could ask for withdrawal from the register at any time-point. All medical images were anonymized and given a code by the Media Department at Sahlgrenska University Hospital before analyses.

Results

Characteristics of the study population

The study population (n = 218) demonstrated mean NRS back 2.90 (SD = 2.71), mean GA 2.13 (SD = 1.19) and mean ODI 21.59 (SD = 19.63). A large spread of the signal intensity measures was observed (*Table 2*) reflecting the heterogenicity of the IVD tissue. In the subregions, the highest mean MSI was observed in ROI 3, which represents the nucleus pulposus. Highest mean value of SD was observed in the most posterior part (ROI 5) of the IVD, the most common location of disc herniation.

Table 2: Mean Signal Intensity (MSI), Standard Deviation of Signal Intensity (SDSI) and SDSI/MSI, of the study population (n=218). The top three rows refer to the hole midsagittal part of the intervertebral disc and the regions of interest (ROI) number 1-5 to the different subregions.

	Mean	Std. Deviation
MSI	112.09	199.49
SDSI	41.67	45.67
SDSI/MSI	0.49	0.13
MSI		
ROI 1	88.27	192.12
ROI 2	116.06	205.55
ROI 3	135.27	221.58
ROI 4	117.47	194.92
ROI 5	100.60	185.05
SDSI		
ROI 1	24.85	26.82
ROI 2	33.94	37.62
ROI 3	37.82	47.90
ROI 4	37.58	44.68
ROI 5	38.75	48.28
SDSI/MSI		
ROI 1	0.45	0.18
ROI 2	0.39	0.13
ROI 3	0.34	0.09
ROI 4	0.40	0.12
ROI 5	0.51	0.16

NRS back outcome groups

Table 3 shows the descriptive statistics of the dichotomized groups with even mean age and level of disc herniation, but with uneven gender distribution, between the groups. The distribution of signal intensity measures (MSI, SDSI and SDSI/MSI) of the whole midsagittal part of the IVDs are displayed in *Figure 6* with similar values and spread in the outcome

groups. No statistically significant difference in signal intensity measures of the whole midsagittal part of the IVDs was found between the outcome groups. Nor in the subregion's ROI 1-5, where no significant difference was found, except for (p=0.045) SDSI in ROI 4 (*Table 4*). ROI 4 represent the border zone between nucleus pulposus and posterior annulus fibrosus (*Figure 5*).

Table 3: Baseline characteristics of successful (Group 0), and unsuccessful (Group 1), surgery outcome groups regarding; Numeric Pain Rating Scale (NRS) back, Global Assessment (GA) and Oswestry Disability Index (ODI).

		NRS back 1-year outcome		GA 1-year outcome		ODI 1-year outcome	
		Group 0 NRS ≤2	Group 1 NRS >2	Group 0 GA ≤2	Group 1 GA>2	Group 0 ODI ≤ 20	Group 1 ODI > 20
Number of Patients		116	102	149	68	120	92
Age	Mean	43	42	43	41	41	43
	SD	16	15	17	14	16	15
Sex	Men (%)	56.0%	39.2%	49.7%	45.6%	55.8%	37.0%
	Woman (%)	44.0%	60.8%	50.3%	54.4%	44.2%	63.0%
Disc level*	L1-L2 (%)	0.9%	0.0%	0.7%	0.0%	0.8%	0.0%
	L2-L3 (%)	0.9%	1.0%	0.7%	1.5%	0.8%	1.1%
	L3-L4 (%)	6.0%	4.9%	7.4%	1.5%	7.5%	3.3%
	L4-L5 (%)	38.8%	45.1%	40.9%	42.6%	36.7%	47.8%
	L5-S1 (%)	53.4%	49.0%	50.3%	54.4%	54.2%	47.8%
BMI**	Mean	26.33	27.23	26.62	27.06	25.71	28.30
	SD	4.48	4.22	3.94	5.45	3.94	4.69
Smokers***	(%)	4.8%	16.7%	5.5%	21.7%	2.3%	18.8%

* Level of disc herniation.
** Body Mass Index (BMI) only reported in n=71 (33%) out of n=218 patients.
*** Smoking habits (at the time of surgery) only reported in n=78 (36%) out of n=218 patients.

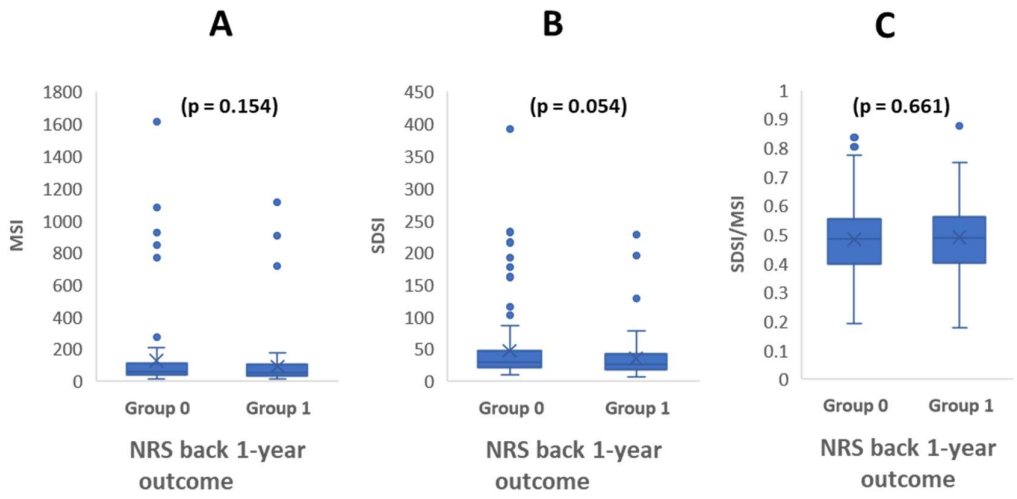

Figure 6: Boxplots showing the distribution of the measures; ***A*** *- Mean signal intensity (MSI),* ***B*** *- Standard deviation of signal intensity (SDSI) and* ***C****- SDSI/MSI, and p-values between successful (Group 0), versus unsuccessful (Group 1), surgery outcome regarding Numeric Pain Rating Scale (NRS) back.*

GA outcome groups

Table 3 shows the descriptive statistics of the dichotomized groups with even mean age and level of disc herniation and similar gender distribution between the groups. The distribution of signal intensity measures (MSI, SDSI and SDSI/MSI) of the whole midsagittal part of the IVDs are displayed in *Figure 7* with similar values and spread in the outcome groups. No statistically significant difference in signal intensity measures of the whole midsagittal part of the IVDs was found between the outcome groups. Nor in the subregion's ROI 1-5, where no significant difference was found (*Table 4*).

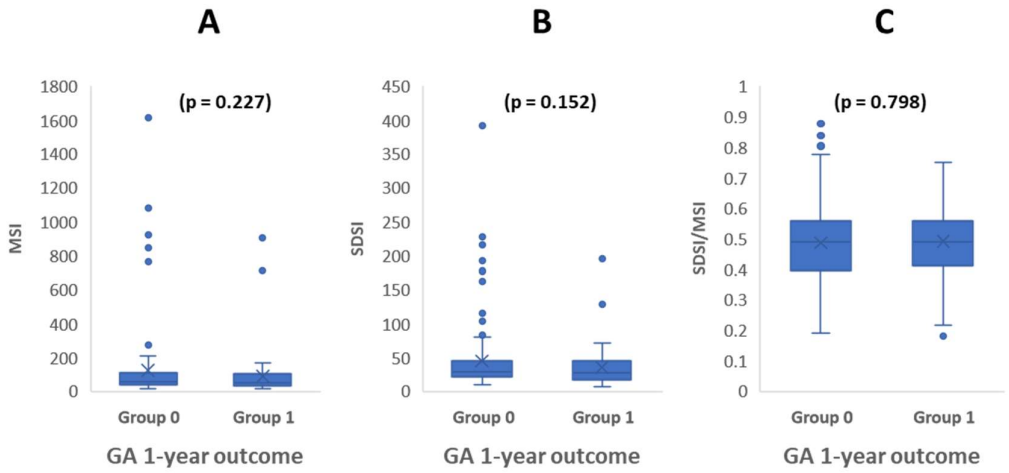

Figure 7: Boxplots showing the distribution of the measures; ***A*** *- Mean signal intensity (MSI),* ***B*** *- Standard deviation of signal intensity (SDSI) and* ***C****- SDSI/MSI, and p-values between successful (Group 0), versus unsuccessful (Group 1), surgery outcome regarding Global Assessment (GA).*

ODI outcome groups

Table 3 shows the descriptive statistics of the dichotomized groups with even mean age and level of disc herniation, but with uneven gender distribution, between the groups. The distribution of signal intensity measures (MSI, SDSI and SDSI/MSI) of the whole midsagittal part of the IVDs are displayed in *Figure 8* with similar values and spread in the outcome groups. No statistically significant difference in signal intensity measures of the whole midsagittal part of the IVDs was found between the outcome groups. Nor in the subregion's ROI 1-5, where no significant difference was found (*Table 4*).

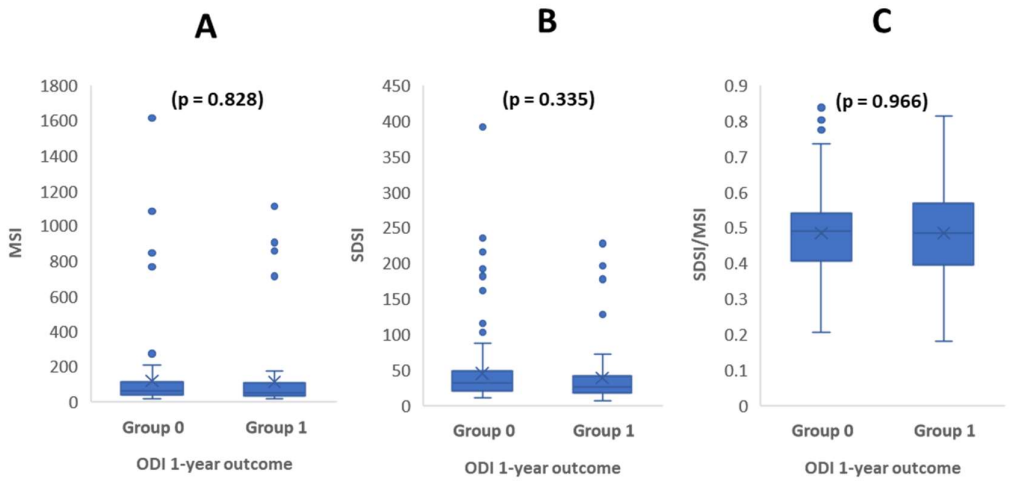

Figure 8: Boxplots showing the distribution of the measures; ***A*** *- Mean signal intensity (MSI),* ***B*** *- Standard deviation of signal intensity (SDSI) and* ***C****- SDSI/MSI, and p-values between successful (Group 0), versus unsuccessful (Group 1), surgery outcome regarding Oswestry Disability Index (ODI)*

Table 4: Group comparison of midsagittal Mean Signal Intensity (MSI), Standard Deviation of Signal Intensity (SDSI) and SDSI/MSI, between successful (Group 0), versus unsuccessful (Group 1), surgery outcome regarding; Numeric Pain Rating Scale (NRS), Global Assessment (GA) and Oswestry Disability Index (ODI). The top three rows refer to the hole midsagittal part of the disc and the region of interest (ROI) number 1-5 to the different subregions.

	NRS back 1-year outcome			GA 1-year outcome			ODI 1-year outcome		
	Group 0 NRS ≤2	Group 1 NRS >2		Group 0 GA ≤2	Group 1 GA>2		Group 0 ODI ≤ 20	Group 1 ODI >20	
	Mean	Mean	p-value	Mean	Mean	p-value	Mean	Mean	p-value
MSI	129.71	92.06	0.154	123.35	87.91	0.227	116.29	110.18	0.828
SDSI	47.08	35.51	0.054	44.73	35.12	0.152	44.68	38.49	0.335
SDSI/MSI	0.48	0.49	0.661	0.49	0.49	0.798	0.48	0.48	0.966
MSI									
ROI1	101.46	73.27	0.281	97.63	68.21	0.297	88.82	90.57	0.948
ROI2	133.39	96.35	0.174	128.25	89.61	0.201	121.18	113.03	0.778
ROI3	157.34	110.16	0.107	148.66	106.61	0.196	142.06	130.50	0.711
ROI4	137.72	94.44	0.092	129.40	91.89	0.190	123.10	113.93	0.738
ROI5	115.83	83.28	0.184	109.15	82.22	0.322	103.62	99.36	0.870
SDSI									
ROI1	27.65	21.67	0.093	26.68	20.75	0.132	26.62	23.00	0.337
ROI2	37.81	29.54	0.096	36.40	28.69	0.162	35.87	31.91	0.455
ROI3	42.95	31.99	0.080	41.14	30.82	0.142	39.79	36.01	0.575
ROI4	43.07	31.33	**0.045**	40.71	30.97	0.137	40.18	34.90	0.402
ROI5	42.82	34.12	0.185	40.36	35.25	0.472	41.10	35.86	0.439
SDSI/MSI									
ROI1	0.45	0.44	0.673	0.46	0.43	0.330	0.46	0.43	0.362
ROI2	0.39	0.40	0.377	0.39	0.40	0.415	0.38	0.40	0.250
ROI3	0.32	0.35	0.061	0.33	0.34	0.687	0.33	0.34	0.434
ROI4	0.39	0.41	0.164	0.40	0.40	0.864	0.39	0.40	0.834
ROI5	0.51	0.52	0.740	0.51	0.51	0.939	0.51	0.50	0.801

Discussion

The aim of this study was to identify IVD characteristics in preoperative MR images and compare them to the 1-year post lumbar disc herniation surgery outcome. This study could not prove any relationship between IVD characteristics and surgery outcome, except regarding SDSI of ROI 4 in the NRS back outcome group where a small significant difference ($p = .045$) were detected. ROI 4 represent the border zone between nucleus pulposus and posterior annulus fibrosus, a possible location of annular tears, which makes this finding interesting. However, no significant difference was observed in the normalized measure SDSI/MSI, in ROI 4, why the strength of evidence in the significant finding is limited. The normalized measure SDSI/MSI, was used to equalize potential differences between scanners for example, due to different signal amplification. However, our findings did not support this.

The results of this study show similar preoperative quantitative measures of disc degeneration, with similar spreading, in patients with different surgical outcome. Our hypothesis was that high level of preoperative disc degeneration, thus low signal intensity measures, would predict worse surgical outcome. This study proved us wrong. However, our method may not be accurate enough due to the limitations of this study (discussed in section *Limitations and Strengths*). More information about the patients, including BMI and smoking habits, needs to be collected to enable statistical adjustments. Thereby, general conclusion of this study cannot be drawn.

The fundamental idea was to identify the IVD characteristics using automatic segmentation based on deep learning. Unfortunately, the automatic segmentation did not perform as desired why semi-automatic segmentation was done. When testing the automatic segmentation, it seemed like the software struggled with separating the IVDs of different levels. Using this method in the analysis would have demanded time spending manual post processing of the images, and the advantage of automatic segmentation had been lost. Automatic segmentation of lumbar MRI has earlier been described as a difficult task (54) and artificial neural networks are computationally advanced and difficult to train (7). In the present study, the training set consisted of only n = 13 manual segmented patients, which might have been too few. The major reason, why the automatic segmentation did not perform as good as desired, was probably the origin of the MRI-series. The MRI examinations in this study were at least made at 15 different clinical department and many different MRI protocols was used. In future work it might be an advantage to supply the software programmers with MRI data from the same scanner model using given MRI protocol.

After rephrasing the research question and switching to another segmentation software, the extracted disc characteristics were limited to signal intensity measures and not disc volume. Disc volume determination requires segmentation of all sagittal MRI slices, often 15-17 slices per patient, a very time-consuming task. The manual segmentation of all MRI slices was made in the trainings-set of n=13 patients. The expenditure of time of this procedure was approximately 1 hour per disc. The lack of time made it impossible to do this procedure on all patients. By comparison, semi-automated segmentation of three mid sagittal slices using our inhouse developed software took approximately 5 minutes per disc. Thus, our in-house semi-automated segmentation software could have been used for volume determination. However,

at the time we realized that the deep-learning software could not fulfill the task the project was near an end.

Methodological considerations

The dichotomization of the outcome measures was made to facilitate the detection of relationship between disc characteristics and the outcomes. In order to get all patient with the same rated outcome in the same group, the dichotomization was performed numerically, thus, not by mean or median. Regarding NRS and ODI, 1-year post surgery scores ≤2 respectively ≤20 are in this study considerer as successful outcome. Regarding GA, the choice of cutoff line can be discussed. Scores 3, meaning the patient is *"somewhat improved"* (*Table 1*), might be considered as a successful or an unsuccessful outcome. However, since a majority (68%) of the patients operated for lumbar disc herniation in 2016 in Swespine, reported GA, regarding back pain, score 1, "*completely pain-free"* (20%), or score 2, "*significantly improved"* (48%) (2), it is reasonable to treat the remaining minority as patients with unsuccessful surgical outcome. In *Table 3,* the baseline characteristic of the three different dichotomized outcomes group are displayed and it was seen that all groups were even in terms of mean age and the level of disc herniation. However, there was an uneven gender distribution, especially when dividing the patients into the NRS and ODI groups. The proportion of women was greater in each group of patients with unsuccessful outcome versus patient with successful outcome, compared to the proportion of men. This was in line with previous studies where it has been shown that the 1 year postoperative outcome is inferior in

woman than in men (67) and further that women are associated with a slower rate of perceived recovery as well as a higher rate of unsatisfactory outcome (68). To test if this skewness of gender distribution might affect the result, the study population was divided in two groups based on gender. Independent t-test was performed for each signal intensity measures. The tests were not associated with any significant difference in none of the measures, nor in none of the subregions (*Appendices 2*). The uneven gender distribution in the outcome groups, therefore, should not affect the results of this study.

Several factors have the potential to affect the 1-year outcome, amongst those are smoking, which is a predictor for unsatisfactory surgical outcome (68, 69). High BMI is, as well, a known risk factor for lumbar disc herniation (70) and overweight is associated with higher recurrence rate after lumbar disc herniation surgery (71), thus, BMI ought to affect the surgery outcome. Unfortunately, only a fraction of the study population (n=218) in this study has reported BMI (n=71, 33%) and smoking habits (n=79, 36%), why statistical adjustment is not adequate.

Limitations and Strengths

This study is limited by the absence of information about what happened to the patients in the time between the time of surgery and the one-year follow-up. There is no information about the surgery itself, how it went, experience of the surgeon and so on. In addition, there is no information about the patients postoperative symptoms until the one year follow-up. Were the

patients following their postoperative recommendation in resting, sick leave and physical rehabilitation training? This can be considered as residual confounders, due to lack of possibility to adjust for these factors.

Another limitation might be the segmentation in the software ITK-SNAP (62) and MATLAB software R206a. It was done by a last year´s medical student with limited experience of medical imaging. If the segmentation had been made by an experienced radiologist or a medical physicist, it might have been more robust and accurate. However, the segmentation was made by one single person, which may be considered as a strength, because else individual difference in the segmentation might have occurred. In order to improve the semi-automatic segmentation, T1-weighted images should have been used for guidance. The border between annulus fibrous and surrounding tissue is more detectable in T1-weighted images, which facilitates the manual segmentation. In the MATLAB-based software, there is a function that transfers and rescales the T1-weighted images to match the corresponding T2-weighted image and thereby enables segmentation on both T1- and T2-weghted images or on the images separately. Unfortunately, the MR images retrieved from the media data base were unsorted and there was no time enough to sort the T1-weighted images to match the T2-weighted images.

Conclusion

Quantitative measures of disc degeneration have the potential to increase the utilization of MRI examinations and to support the development of automatic analysis in medical imaging. In this study IVD characteristics from preoperative MR images where identified using semi-automated segmentation. However, this method could not find any relationship, between MRI characteristics preoperatively for the disc and the surgical outcome measured by PROM´s one year postoperatively. Disc degeneration is linked to low back pain and might be the reason why some patients do not experience backpain relief after lumbar disc herniation surgery. To show this, further studies are encouraged using standardized MRI protocol and scanners, and more patient data enabling adjustment of confounders.

Populärvetenskaplig sammanfattning

Segmentering av ryggdiskar – en jämförelse av preoperativa MR bilder med resultatet 1 år efter diskbråcksoperation

Diskbråck är en vanlig åkomma som skapar problem med framförallt smärta i benen för den drabbade patienten. Det beror på att en disk mellan ryggradens kotor kollapsar och trycker på den bakomliggande ryggmärgen. Diskbråck kan behandlas med kirurgi där det som trycker på ryggmärgen tas bort. Inför en eventuell operation undersöks ryggen med en s.k. magnetkamera (MR) som ger en bild av disken. Bilden används för ställningstagande till om diskbråcket ska opereras eller ej.

Operation av diskbråck förbättrar ofta patienternas bensmärta avsevärt men tyvärr kvarstår problem med ryggvärk hos vissa patienter, möjligen kopplat till grad av diskdegeneration (åldrande). Denna degeneration kan mätas med hjälp av MR, redan före operation. Studien syftar till att undersöka diskars utseende på MR bilder innan diskbråcksoperation och jämföra dem med utfallet 1 år efter operation.

218st diskbråcksopererade patienter hämtades ur det svenska ryggregistret. Patienternas MR bilder från före operationen togs fram och den sjuka disken på varje patient studerades. Detta gjordes genom segmentering av diskens mittersta del, vilket betyder att disken ringas in i MR bilden. Diskens utseende i form av signalintensitet analyseras, d.v.s. hur starkt disken lyser i bilden och skillnaden av denna signal mellan olika områden av disken beräknades. Det är ett mått på hur disken mår och graden av dess degeneration. Därefter jämfördes om det fanns en

skillnad i signalintensitet i den sjuka disken mellan patienter med lyckat resp. mindre lyckat, operationsresultat.

Den statistiska analysen visar att det inte är någon statistisk skillnad i utseende (signalintensitet) av den sjuka disken mellan patienter med lyckat respektive mindre lyckat operationsresultat. Studien kan därför inte påvisa någon koppling mellan den sjuka diskens utseende före operation och resultatet 1 år efter operation. Det finns många faktorer, oberoende av diskens utseende, som påverkar operationsresultatet och som i studien inte tagits hänsyn till. Uppgifter om hur själva operationen gick, kirurgens erfarenhet och hur patienten skött sin rehabilitering hade varit önskvärt för att kunna justera för dessa faktorer. I denna studie har även bilder från olika MR apparater med olika inställningar använts. Det är svårt att avgöra hur detta påverkat resultatet. I fortsatta studier bör hänsyn tas till detta och val av MR apparater och inställningar bör standardiseras.

Acknowledgement

I would like to thank my supervisors Helena Brisby, Kerstin Lagerstrand and Olof Westin for all support and guidance through this project. Also, thanks to Christian Waldenberg, Evin Papalini, Sathiesh Kaliyugara, Alexander Selvikvåg Lundervold and Joel Beck for calculations, “segmenting education”, programming, and database collection.

References

1. Jordan J, Konstantinou K, O'Dowd J. Herniated lumbar disc. BMJ Clin Evid. 2011;2011.
2. Fritzell P HO, Gerdhem P, Abbott A, Songsong A, Parai C, et al. 2018 Anual report - Fullow up of spine surgery performed in Sweden in 2017. SWEDISH SOCIETY OF SPINAL SURGEONS. 2018.
3. Petersen T, Laslett M, Juhl C. Clinical classification in low back pain: best-evidence diagnostic rules based on systematic reviews. BMC Musculoskelet Disord. 2017;18(1):188.
4. Janssen ME, Bertrand SL, Joe C, Levine MI. Lumbar herniated disk disease: comparison of MRI, myelography, and post-myelographic CT scan with surgical findings. Orthopedics. 1994;17(2):121-7.
5. Brinjikji W, Luetmer PH, Comstock B, Bresnahan BW, Chen LE, Deyo RA, et al. Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. AJNR Am J Neuroradiol. 2015;36(4):811-6.
6. Leven D, Passias PG, Errico TJ, Lafage V, Bianco K, Lee A, et al. Risk Factors for Reoperation in Patients Treated Surgically for Intervertebral Disc Herniation: A Subanalysis of Eight-Year SPORT Data. J Bone Joint Surg Am. 2015;97(16):1316-25.
7. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys. 2019;29(2):102-27.
8. Mazurowski MA, Buda M, Saha A, Bashir MR. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. J Magn Reson Imaging. 2019;49(4):939-54.
9. Jamaludin A, Lootus M, Kadir T, Zisserman A, Urban J, Battie MC, et al. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: Automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. Eur Spine J. 2017;26(5):1374-83.
10. Moore KL. Clinically oriented anatomy. Eighth edition. ed. Agur AMR, Dalley AF, editors: Philadelphia : Wolters Kluwer; 2018.
11. Raj PP. Intervertebral disc: anatomy-physiology-pathophysiology-treatment. Pain Pract. 2008;8(1):18-44.
12. Bogduk N. Clinical anatomy of the lumbar spine and sacrum. 4. ed. ed. Edinburgh: Edinburgh : Church Livingstone; 2005.
13. Moore RJ. The vertebral endplate: disc degeneration, disc regeneration. Eur Spine J. 2006;15 Suppl 3:S333-7.
14. Boos N, Weissbach S, Rohrbach H, Weiler C, Spratt KF, Nerlich AG. Classification of age-related changes in lumbar intervertebral discs: 2002 Volvo Award in basic science. Spine (Phila Pa 1976). 2002;27(23):2631-44.
15. Adams MA, Lama P, Zehra U, Dolan P. Why do some intervertebral discs degenerate, when others (in the same spine) do not? Clin Anat. 2015;28(2):195-204.
16. Buckwalter JA. Aging and degeneration of the human intervertebral disc. Spine (Phila Pa 1976). 1995;20(11):1307-14.
17. Osti OL, Vernon-Roberts B, Moore R, Fraser RD. Annular tears and disc degeneration in the lumbar spine. A post-mortem study of 135 discs. J Bone Joint Surg Br. 1992;74(5):678-82.
18. Li L, Zhou Z, Xiong W, Fang J, Li Y, Jiao Z, et al. Characterization of the microstructure of the intervertebral disc in patients with chronic low back pain by diffusion kurtosis imaging. Eur Spine J. 2019;28(11):2517-25.

19. Fardon DF, Williams AL, Dohring EJ, Murtagh FR, Gabriel Rothman SL, Sze GK. Lumbar disc nomenclature: version 2.0: Recommendations of the combined task forces of the North American Spine Society, the American Society of Spine Radiology and the American Society of Neuroradiology. Spine J. 2014;14(11):2525-45.
20. Rajasekaran S, Bajaj N, Tubaki V, Kanna RM, Shetty AP. ISSLS Prize winner: The anatomy of failure in lumbar disc herniation: an in vivo, multimodal, prospective study of 181 subjects. Spine (Phila Pa 1976). 2013;38(17):1491-500.
21. Deyo RA, Mirza SK. CLINICAL PRACTICE. Herniated Lumbar Intervertebral Disk. N Engl J Med. 2016;374(18):1763-72.
22. Amin RM, Andrade NS, Neuman BJ. Lumbar Disc Herniation. Curr Rev Musculoskelet Med. 2017;10(4):507-16.
23. Stromqvist F, Stromqvist B, Jonsson B, Karlsson MK. Surgical treatment of lumbar disc herniation in different ages-evaluation of 11,237 patients. Spine J. 2017;17(11):1577-85.
24. Peul WC, van Houwelingen HC, van den Hout WB, Brand R, Eekhof JA, Tans JT, et al. Surgery versus prolonged conservative treatment for sciatica. N Engl J Med. 2007;356(22):2245-56.
25. Wade KR, Robertson PA, Thambyah A, Broom ND. How healthy discs herniate: a biomechanical and microstructural study investigating the combined effects of compression rate and flexion. Spine (Phila Pa 1976). 2014;39(13):1018-28.
26. Nachemson AL. Disc pressure measurements. Spine (Phila Pa 1976). 1981;6(1):93-7.
27. Gibson JN, Waddell G. Surgical interventions for lumbar disc prolapse: updated Cochrane Review. Spine (Phila Pa 1976). 2007;32(16):1735-47.
28. Kreiner DS, Hwang SW, Easa JE, Resnick DK, BaiSDSIen JL, Bess S, et al. An evidence-based clinical guideline for the diagnosis and treatment of lumbar disc herniation with radiculopathy. Spine J. 2014;14(1):180-91.
29. Matsumoto M, Okada E, Ichihara D, Watanabe K, Chiba K, Toyama Y, et al. Age-related changes of thoracic and cervical intervertebral discs in asymptomatic subjects. Spine (Phila Pa 1976). 2010;35(14):1359-64.
30. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS. Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med. 1994;331(2):69-73.
31. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. Jama. 2006;296(20):2441-50.
32. Chedid KJ, Chedid MK. The "tract" of history in the treatment of lumbar degenerative disc disease. Neurosurg Focus. 2004;16(1):E7.
33. Parisien RC, Ball PA. William Jason Mixter (1880-1958). Ushering in the "dynasty of the disc". Spine (Phila Pa 1976). 1998;23(21):2363-6.
34. Mixter WJ BJ. Rupture of the intervertebral disc with involvement of the spinal canal. New Engl J Med. 1934;1934;211:210—5.
35. Blamoutier A. Surgical discectomy for lumbar disc herniation: surgical techniques. Orthop Traumatol Surg Res. 2013;99(1 Suppl):S187-96.
36. Koebbe CJ, Maroon JC, Abla A, El-Kadi H, Bost J. Lumbar microdiscectomy: a historical perspective and current technical considerations. Neurosurg Focus. 2002;13(2):E3.
37. Rasouli MR, Rahimi-Movaghar V, Shokraneh F, Moradi-Lakeh M, Chou R. Minimally invasive discectomy versus microdiscectomy/open discectomy for symptomatic lumbar disc herniation. Cochrane Database Syst Rev. 2014(9):Cd010328.
38. Choi KC, Kim JS, Park CK. Percutaneous Endoscopic Lumbar Discectomy as an Alternative to Open Lumbar Microdiscectomy for Large Lumbar Disc Herniation. Pain Physician. 2016;19(2):E291-300.

39. Chen Z, Zhang L, Dong J, Xie P, Liu B, Wang Q, et al. Percutaneous transforaminal endoscopic discectomy compared with microendoscopic discectomy for lumbar disc herniation: 1-year results of an ongoing randomized controlled trial. J Neurosurg Spine. 2018;28(3):300-10.
40. Alvi MA, Kerezoudis P, Wahood W, Goyal A, Bydon M. Operative Approaches for Lumbar Disc Herniation: A Systematic Review and Multiple Treatment Meta-Analysis of Conventional and Minimally Invasive Surgeries. World Neurosurg. 2018;114:391-407.e2.
41. Kamper SJ, Ostelo RW, Rubinstein SM, Nellensteijn JM, Peul WC, Arts MP, et al. Minimally invasive surgery for lumbar disc herniation: a systematic review and meta-analysis. Eur Spine J. 2014;23(5):1021-43.
42. Ruetten S, Komp M, Merk H, Godolias G. Full-endoscopic interlaminar and transforaminal lumbar discectomy versus conventional microsurgical technique: a prospective, randomized, controlled study. Spine (Phila Pa 1976). 2008;33(9):931-9.
43. Grover VP, Tognarelli JM, Crossey MM, Cox IJ, Taylor-Robinson SDSI, McPhail MJ. Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians. J Clin Exp Hepatol. 2015;5(3):246-55.
44. Rockall AG. Diagnostic Imaging. 7. ed. ed. Hatrick A, Armstrong P, Wastie M, editors. Chichester: Chichester : Wiley-Blackwell; 2013.
45. Messner A, Stelzeneder D, Trattnig S, Welsch GH, Schinhan M, Apprich S, et al. Does T2 mapping of the posterior annulus fibrosus indicate the presence of lumbar intervertebral disc herniation? A 3.0 Tesla magnetic resonance study. Eur Spine J. 2017;26(3):877-83.
46. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine (Phila Pa 1976). 2001;26(17):1873-8.
47. Luoma K, Vehmas T, Riihimaki H, Raininko R. Disc height and signal intensity of the nucleus pulposus on magnetic resonance imaging as indicators of lumbar disc degeneration. Spine (Phila Pa 1976). 2001;26(6):680-6.
48. Hu X, Chen M, Pan J, Liang L, Wang Y. Is it appropriate to measure age-related lumbar disc degeneration on the mid-sagittal MR image? A quantitative image study. Eur Spine J. 2018;27(5):1073-81.
49. Benneker LM, Heini PF, Anderson SE, Alini M, Ito K. Correlation of radiographic and MRI parameters to morphological and biochemical assessment of intervertebral disc degeneration. Eur Spine J. 2005;14(1):27-35.
50. Bote-Curiel L, Muñoz-Romero S, Gerrero-Curieses A, Rojo-Álvarez LJ. Deep Learning and Big Data in Healthcare: A Double Review for Critical Beginners. Applied Sciences. 2019;9(11).
51. Hinton G. Deep Learning-A Technology With the Potential to Transform Health Care. Jama. 2018;320(11):1101-2.
52. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng. 2018;2(3):158-64.
53. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-8.
54. Ghosh S, Chaudhary V. Supervised methods for detection and segmentation of tissues in clinical lumbar MRI. Comput Med Imaging Graph. 2014;38(7):639-49.
55. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017;60(6):84-90.
56. Price DD, Bush FM, Long S, Harkins SW. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. Pain. 1994;56(2):217-26.
57. Chapman JR, Norvell DC, Hermsmeyer JT, Bransford RJ, DeVine J, McGirt MJ, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. Spine (Phila Pa 1976). 2011;36(21 Suppl):S54-68.

58. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. Spine (Phila Pa 1976). 2005;30(11):1331-4.
59. Parai C, Hagg O, Lind B, Brisby H. The value of patient global assessment in lumbar spine surgery: an evaluation based on more than 90,000 patients. Eur Spine J. 2018;27(3):554-63.
60. Hagg O, Fritzell P, Oden A, Nordwall A. Simplifying outcome measurement: evaluation of instruments for measuring outcome after fusion surgery for chronic low back pain. Spine (Phila Pa 1976). 2002;27(11):1213-22.
61. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976). 2000;25(22):2940-52; discussion 52.
62. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage. 2006;31(3):1116-28.
63. Videman T, Gibbons LE, Battie MC. Age- and pathology-specific measures of disc degeneration. Spine (Phila Pa 1976). 2008;33(25):2781-8.
64. Waldenberg C, Hebelka H, Brisby H, Lagerstrand KM. MRI histogram analysis enables objective and continuous classification of intervertebral disc degeneration. Eur Spine J. 2018;27(5):1042-8.
65. Waldenberg C, Hebelka H, Brisby H, Lagerstrand KM. Differences in IVD characteristics between low back pain patients and controls associated with HIZ as revealed with quantitative MRI. PLoS One. 2019;14(8):e0220952.
66. Nilsson M, Lagerstrand K, Kasperska I, Brisby H, Hebelka H. Axial loading during MRI influences T2-mapping values of lumbar discs: a feasibility study on patients with low back pain. Eur Spine J. 2016;25(9):2856-63.
67. Stromqvist F, Stromqvist B, Jonsson B, Karlsson MK. Inferior Outcome of Lumbar Disc Surgery in Women Due to Inferior Preoperative Status: A Prospective Study in 11,237 Patients. Spine (Phila Pa 1976). 2016;41(15):1247-52.
68. Peul WC, Brand R, Thomeer RT, Koes BW. Influence of gender and other prognostic factors on outcome of sciatica. Pain. 2008;138(1):180-91.
69. Kerr D, Zhao W, Lurie JD. What Are Long-term Predictors of Outcomes for Lumbar Disc Herniation? A Randomized and Observational Study. Clin Orthop Relat Res. 2015;473(6):1920-30.
70. Bostman OM. Body mass index and height in patients requiring surgery for lumbar intervertebral disc herniation. Spine (Phila Pa 1976). 1993;18(7):851-4.
71. Fotakopoulos G, Makris D, Kotlia P, Tzerefos C, Fountas K. Recurrence Is Associated With Body Mass Index in Patients Undergoing a Single-Level Lumbar Disc Herniation Surgery. J Clin Med Res. 2018;10(6):486-92.

Appendices

Appendices 1: Oswestry Low Back Pain Disability Questionnaire. For each section the total possible score is 5: if the first statement is marked the section score = 0; if the last statement is marked, it = 5. (http://www.rehab.msu.edu/_files/_docs/oswestry_low_back_disability.pdf)

Oswestry Low Back Pain Disability Questionnaire

Instructions

This questionnaire has been designed to give us information as to how your back or leg pain is affecting your ability to manage in everyday life. Please answer by checking ONE box in each section for the statement which best applies to you. We realise you may consider that two or more statements in any one section apply but please just shade out the spot that indicates the statement which most clearly describes your problem.

Section 1 – Pain intensity

- ☐ I have no pain at the moment
- ☐ The pain is very mild at the moment
- ☐ The pain is moderate at the moment
- ☐ The pain is fairly severe at the moment
- ☐ The pain is very severe at the moment
- ☐ The pain is the worst imaginable at the moment

Section 2 – Personal care (washing, dressing etc)

- ☐ I can look after myself normally without causing extra pain
- ☐ I can look after myself normally but it causes extra pain
- ☐ It is painful to look after myself and I am slow and careful
- ☐ I need some help but manage most of my personal care
- ☐ I need help every day in most aspects of self-care
- ☐ I do not get dressed, I wash with difficulty and stay in bed

Section 3 – Lifting

- ☐ I can lift heavy weights without extra pain
- ☐ I can lift heavy weights but it gives extra pain
- ☐ Pain prevents me from lifting heavy weights off the floor, but I can manage if they are conveniently placed eg. on a table
- ☐ Pain prevents me from lifting heavy weights, but I can manage light to medium weights if they are conveniently positioned
- ☐ I can lift very light weights
- ☐ I cannot lift or carry anything at all

Section 4 – Walking*

- ☐ Pain does not prevent me walking any distance
- ☐ Pain prevents me from walking more than 1 mile
- ☐ Pain prevents me from walking more than 1/2 mile
- ☐ Pain prevents me from walking more than 100 yards
- ☐ I can only walk using a stick or crutches
- ☐ I am in bed most of the time

Section 5 – Sitting

- ☐ I can sit in any chair as long as I like
- ☐ I can only sit in my favourite chair as long as I like
- ☐ Pain prevents me sitting more than one hour
- ☐ Pain prevents me from sitting more than 30 minutes
- ☐ Pain prevents me from sitting more than 10 minutes
- ☐ Pain prevents me from sitting at all

Section 6 – Standing

- ☐ I can stand as long as I want without extra pain
- ☐ I can stand as long as I want but it gives me extra pain
- ☐ Pain prevents me from standing for more than 1 hour
- ☐ Pain prevents me from standing for more than 30 minutes
- ☐ Pain prevents me from standing for more than 10 minutes
- ☐ Pain prevents me from standing at all

Section 7 – Sleeping

- ☐ My sleep is never disturbed by pain
- ☐ My sleep is occasionally disturbed by pain
- ☐ Because of pain I have less than 6 hours sleep
- ☐ Because of pain I have less than 4 hours sleep
- ☐ Because of pain I have less than 2 hours sleep
- ☐ Pain prevents me from sleeping at all

Section 8 – Sex life (if applicable)

- ☐ My sex life is normal and causes no extra pain
- ☐ My sex life is normal but causes some extra pain
- ☐ My sex life is nearly normal but is very painful
- ☐ My sex life is severely restricted by pain
- ☐ My sex life is nearly absent because of pain
- ☐ Pain prevents any sex life at all

Section 9 – Social life

- ☐ My social life is normal and gives me no extra pain
- ☐ My social life is normal but increases the degree of pain
- ☐ Pain has no significant effect on my social life apart from limiting my more energetic interests eg, sport
- ☐ Pain has restricted my social life and I do not go out as often
- ☐ Pain has restricted my social life to my home
- ☐ I have no social life because of pain

Section 10 – Travelling

- ☐ I can travel anywhere without pain
- ☐ I can travel anywhere but it gives me extra pain
- ☐ Pain is bad but I manage journeys over two hours
- ☐ Pain restricts me to journeys of less than one hour
- ☐ Pain restricts me to short necessary journeys under 30 minutes
- ☐ Pain prevents me from travelling except to receive treatment

Appendices 2: Comparison of midsagittal Mean Signal Intensity (MSI), Standard Deviation of Signal Intensity (SDSI) and SDSI/MSI, in the study population divided by gender. The top three rows refer to the hole midsagittal part of the disc and the regions of interest (ROI) number 1-5 to the different subregions.

	Gender		
	Men	Woman	
	Mean	Mean	p-value
MSI	104.63	119.03	0.595
SDSI	40.28	42.95	0.667
SDSI/MSI	0.49	0.48	0.754
MSI			
ROI 1	81.55	94.50	0.620
ROI 2	108.09	123.47	0.582
ROI 3	126.17	143.72	0.560
ROI 4	111.41	123.10	0.659
ROI 5	94.52	106.25	0.641
SDSI			
ROI 1	24.04	25.60	0.668
ROI 2	32.68	35.11	0.635
ROI 3	35.59	39.90	0.508
ROI 4	36.58	38.50	0.752
ROI 5	38.21	39.25	0.874
SDSI/MSI			
ROI 1	0.45	0.45	0.895
ROI 2	0.39	0.40	0.441
ROI 3	0.33	0.34	0.738
ROI 4	0.40	0.39	0.787
ROI 5	0.52	0.51	0.601