Title

Adjacent disc degeneration after lumbar total disc replacement or non-operative treatment: A randomized study with eight-year follow-up.

Device status (statement)

The device (Prodisc II) is FDA approved for this indication (suspected discogenic low back pain).

Names, affiliations and addresses of the authors

Håvard Furunes
Innlandet Hospital Gjøvik, Department of Surgery, Innlandet Hospital Gjøvik, Kyrre Grepps gate 11, 2819   GJØVIK, Norway
University of Oslo, Postbox 1072 Blindern, 0316 OSLO, Norway
Oslo University Hospital Ullevål, Research- and comunication unit for musculoskeletal health (FORMI), Building 37B, Postbox 4950 Nydalen, 0424 OSLO, Norway
E-mail: havardfurunes@gmail.com

Christian Hellum
Division of Orthopaedic Surgery, Oslo University Hospital, Postbox 4950 Nydalen, 0424 OSLO, Norway
E-mail: CHRHEL@ous-hf.no

Ansgar Espeland
Department of Radiology, Haukeland University Hospital, Jonas Liesvei 65, 5021 Bergen, Norway.
Department of Clinical Medicine, University of Bergen, Postbox 7804, 5020 Bergen, Norway
E-mail: ansgar.espeland@gmail.com

Jens Ivar Brox
University of Oslo, Postbox 1072 Blindern, 0316 OSLO, Norway
Department for Physical Medicine and Rehabilitation, Oslo University Hospital, Postbox 4950 Nydalen, 0424 OSLO, Norway
E-mail: jbrox@ous-hf.no

Milada Cvancarova Småstuen
Faculty of Health Sciences, Oslo Metropolitan University, Postbox 4 St. Olavs Plass, 0130 OSLO, Norway
Oslo University Hospital Ullevål, Research- and comunication unit for musculoskeletal health (FORMI), Building 37B, Postbox 4950 Nydalen, 0424 OSLO, Norway
E-mail: miladacv@medisin.uio.no

Linda Berg
Department of Radiology, Nordland Hospital, Postbox 1480, 8092 Bodø, Norway
Department of Clinical Medicine, Faculty of Health Sciences, University of Tromsø, 9037 Tromsø, Norway
E-mail: linda.berg@nlsh.no

Kjersti Storheim

Oslo University Hospital Ullevål, Research- and comunication unit for musculoskeletal health (FORMI), Building 37B, Postbox 4950 Nydalen, 0424 OSLO, Norway
E-mail: kjersti.storheim@medisin.uio.no

E-mail and phone number of the corresponding author

Email: *havardfurunes@gmail.com*
Phone number: 0047 99041763

Adjacent disc degeneration after lumbar total disc replacement or non-operative treatment: A randomized study with eight-year follow-up

Abstract

Study design. Randomized controlled multicenter trial with eight-year follow-up.

Objective. To assess the long-term development of adjacent disc degeneration (ADD) after lumbar total disc replacement (TDR) or non-operative treatment, and to analyze the association between ADD development and clinical outcome.

Summary of background data. TDR was introduced as a motion-preserving alternative to spinal fusion, which has been reported to increase the risk of ADD. However, ADD may develop naturally regardless of any surgery, and no randomized study has assessed the long-term development of ADD after TDR versus non-operative treatment.

Methods. The study included 126 of the 173 patients with chronic low back pain (LBP) originally included in a randomized study comparing TDR with multidisciplinary rehabilitation. Magnetic resonance imaging (MRI) of the lumbar spine was performed before treatment and at eight-year follow-up. ADD was categorized as increased or not increased based on an evaluation of Modic changes, disc height reduction, disc contour, herniation size, nucleus pulposus signal and posterior high intensity zones. We used a $\chi 2$ test or a Fisher's exact test to compare crude proportions, and multiple linear regressions to analyze the association between increased ADD (yes/no) and change in Oswestry Disability Index (ODI) from pre-treatment to follow-up.

Results. ADD increased (for at least one ADD variable) in 23 of 57 patients (40%) treated non-operatively, and 29 of 69 patients (42%) treated with TDR (p=0.86). We found no significant associations between ADD increase and the change in ODI.

Conclusions. Increased ADD occurred with similar frequency after TDR and after non-operative treatment, and was not related to the clinical outcome at eight-year follow-up.

Key words: Low back pain, total disc replacement, non-operative treatment, adjacent disc degeneration.

Level of evidence: 1

Key points:

- No differences were observed between lumbar total disc replacement and non-operative treatment in the development of adjacent disc degeneration at eight-year follow-up.
- No association was found between increased adjacent disc degeneration and the clinical outcome.
- Larger studies with longer follow-up are warranted.

Mini abstract:

In this eight-year follow-up of a randomized multicenter study comparing lumbar total disc replacement with non-operative treatment, the development of adjacent disc degeneration did not differ between the treatment groups and was not related to clinical outcome.

Introduction

Concern about adjacent disc degeneration (ADD) following spinal fusion has contributed to the development and implementation of motion preserving alternatives such as lumbar total disc replacement (TDR) in the surgical treatment of low back pain (LBP) [1 2]. According to the literature, ADD is mainly influenced by aging and genetics [3], and may develop regardless of any surgery [1 4 5]. In a meta-analysis [6] of four randomized studies with a mean follow-up of 13 years after spinal fusion or non-operative treatment, ADD was more strongly related to aging than to fusion, but was not associated with the clinical outcome. In the only randomized study of ADD after TDR versus non-operative treatment [7], ADD was neither related to TDR nor the clinical outcome at two-year follow-up. The aim of the present study was to assess the long-term development of ADD after TDR or non-operative treatment, and to analyze the association between ADD development and the clinical outcome.

Materials and Methods

This is a randomized multicenter study with eight-year follow-up conducted at five university hospitals in Norway [8]. The eight-year follow-up was approved by the Norwegian Regional Ethical Committee South East C (2011/2177). The project was conducted in accordance with the Helsinki Declaration and the ICH-GCP guidelines and it was registered at www.clinicaltrial.gov under the identifier NCT01704677 before it commenced. Results are reported according to the CONSORT standard.

Patients

Inclusion criteria for the original randomized trial were age 25-55 years, LBP as the main symptom for at least one year, Oswestry Disability Index (ODI) score ≥ 30 and degenerative changes at the disc levels L4/L5 and/or L5/S1 on magnetic resonance imaging (MRI). For further details see Hellum et al [9]. The present study included patients available for both

radiological and clinical examination at eight-year follow-up. Patients who had been operated with lumbar fusion during follow-up were excluded.

Interventions

Non-operative treatment

The non-operative treatment consisted of modern multidisciplinary rehabilitation with a cognitive approach and supervised physical exercise over three to five weeks, according to Brox et al [10].

TDR

In the TDR procedure, the degenerative intervertebral discs L4/L5 and/or L5/S1 were removed and replaced with artificial discs (ProDisc II, Synthes Spine), as detailed previously [9].

Clinical outcome

Pain and disability were evaluated with the Norwegian version 2.0 of the ODI [11 12] (scores range from 0 to 100, with a lower score indicating less pain and disability). LBP was measured with a visual analogue scale (VAS) ranging from 0 (no pain) to 100 (worst pain imaginable). We also collected information on reoperations due to ADD.

MRI evaluation

MRI variables for evaluating ADD were analyzed at the nearest level above the implanted or degenerated index level, i.e. at L3/L4 or L4/L5. MRI performed before treatment and at eight-year follow-up included (a) sagittal T2-weighted fast spin echo (FSE) and/or DRIVE images (FSE with 90° flip-back pulse); (b) sagittal T1-weighted spin echo or fast fluid-attenuated inversion-recovery images; and (c) axial images of the lower lumbar levels (T2-, T1- or proton density-weighted). All follow-up MRIs and > 90% of pre-treatment MRIs were from 1.5 T units. Metal artefact reducing techniques [13] were used on 97% of follow-up MRIs and implied thin slices (3 mm) and increased pixel bandwidth (BW), echo train length (ETL) and

number of excitations (NEX). These follow-up MRIs included (a) sagittal T2-weighted FSE images: repetition time (TR) / echo time (TE), 4480-5000 ms / 92-94 ms; matrix 448x448; BW 413; ETL 30; NEX 3, (b) sagittal T1-weighted FSE images: TR / TE, 549-610 ms / 8 ms; matrix 448x448; BW 698; ETL 7; NEX 4, and (c) axial T2-weighted FSE images: TR / TE, 4201-5190 ms / 72-92 ms; matrix 320x320 or 512x512; BW 390-413; ETL 20; NEX 2-3.

Pre-treatment and follow-up images were anonymized, presented together in random order and evaluated independently by two radiologists from different institutions who had more than 15 years of experience in spine imaging. The observers could not be blinded to treatment, but were blinded to clinical data. They rated changes in MRI findings by comparing eight-year follow-up and pre-treatment images on a clinical picture archiving and communication system (PACS) unit.

Six ADD variables were rated: Extent of Modic changes [14 15], disc height [16], disc contour [17], disc herniation size [17], nucleus pulposus signal [18] and posterior high intensity zone (HIZ) [19] (Table 1). Conclusive ADD ratings were based on both the observers' independent ratings and consensus between them in every instance of disagreement (review and discussion of images and findings). The interobserver agreement on increased rating value (yes/no) from pre-treatment to follow-up was fair for disc contour (kappa 0.37, disagreement 21%, mean yes rate 21%) and mostly good for the five other ADD variables (kappa 0.60-0.76, disagreement 3-10%, mean yes rate of 27% for nucleus pulposus signal and otherwise 7-12%). [20].

Statistical analysis

The primary outcome measure was increased ADD (yes/no), defined as increased rating value from pre-treatment to follow-up for at least one ADD variable, as in a previous report [7]. Possible rating values are defined in Table 1. Decreased ADD (yes/no) was defined as decreased rating value for at least one ADD variable. Unchanged ADD implied that all rating

values were unchanged. Two patients who had an increased rating value for one ADD variable and a decreased rating value for a different ADD variable were both classified as having increased ADD.

For each ADD variable, we compared changes in ratings between the treatment groups. We performed crude comparisons using a χ2 test or a Fisher's exact test for categorical variables and an independent two-sided t test for continuous variables. Patients who were randomized to TDR but not operated were analyzed in the non-operative group, and patients randomized to rehabilitation who later received TDR were analyzed in the TDR group, according to as-treated principles (Figure 1).

In a sensitivity analysis, we compared the proportions of patients with overall increased ADD in each group after excluding patients randomized to rehabilitation who later received TDR.

In order to analyze possible associations between increased ADD and the clinical outcome adjusted for possible confounders, we fitted a multiple linear regression model. In this model, we excluded those who were not treated according to randomization, since the treatments influence the clinical outcome. The model included ODI change from baseline to follow-up as the dependent variable and the following independent variables: Developed / increased extent of Modic changes (yes/no), disc height reduction (yes/no), disc contour worsening (yes/no), decreased nucleus pulposus signal (yes/no), developed HIZ (yes/no), age, gender and type of treatment (non-operative / TDR). Increased herniation size occurred in only one patient and was therefore not part of the regression model. The model fit was good, with normally distributed residuals. The results are presented as an estimate of beta with 95% confidence intervals (CI). We also performed a multiple logistic regression with the same independent variables as above and "a satisfactory symptom state" (ODI ≤ 22 points at follow-up) (yes/no) [21] as the dependent variable. A significance level of 5% was used for all analyses. All

analyses were considered exploratory so no correction for multiple testing was done. All statistical analyses were performed using SPSS version 24.0.

Post hoc analysis

Post hoc, we analyzed the severity of ADD development based on the number of worsened ADD variables. We compared the proportions of patients from each treatment group with increased rating for none, one, two, three, four, five and six ADD variables.

Results

In the original TDR group, 77 of 86 patients received the TDR, 14 were not available for follow-up, four patients could not be analyzed since they were operated on with spinal fusion, and in one patient, the pre-treatment MRI was untraceable. Therefore, only 58 of the original TDR group could be included. In addition, we included 11 patients randomized to rehabilitation who crossed over and were treated with TDR (median time since surgery was 74 (range 61-85) months). Consequently, 69 patients treated with TDR were analyzed. In the original rehabilitation group, 14 of 87 patients had been treated with TDR and five with spinal fusion, 15 were not available for follow-up and one patient had an untraceable pre-treatment MRI. Therefore, only 52 of the original rehabilitation group could be included in the analyses. In addition, we included five patients randomized to surgery who were not operated. Thus, 57 patients treated non-operatively were analyzed (Figure 1). The two treatment groups had similar pre-treatment clinical, demographical and radiological characteristics (Table 2). The 126 patients included in the present analyses had similar pre-treatment clinical, demographical and radiological characteristics, and similar outcome measures at eight-year follow-up, as did the 47 patients who could not be included ($p \geq 0.12$ for all analyzed variables, data not shown).

At eight-year follow-up, 23 patients (40%) in the non-operative group and 29 patients (42%) in the TDR group had increased ADD (p=0.86). The increase was due to developed / increased extent of Modic changes, development of HIZ, decreased nucleus pulposus signal and/or disc height, worsened disc contour, and/or increased herniation size (Figure 2, Table 3). Three patients (5%) treated non-operatively versus two patients (3%) treated with TDR had decreased ADD (p=0.66). Regression of ADD was due to disappearance of HIZ in three patients, disappearance of Modic changes in one patient and regression of disc herniation in one patient. The change in rating from pre-treatment to eight-year follow-up did not differ significantly between the treatment groups for any of the ADD variables (Table 3).

The sensitivity analysis confirmed our results. After exclusion of the 11 patients randomized to rehabilitation and treated with TDR, 23 patients (40%) in the non-operative group and 24 patients (41%) in the TDR group had increased ADD (p=0.89).

None of the patients who were originally treated non-operatively had been operated at the adjacent level. One patient treated with TDR at L5/S1 had been re-operated 17 months after the initial procedure with fusion from L4 to S1, thus fusing the adjacent level.

In the multiple linear regression analysis of the association between increased ADD and the clinical outcome, the only variable that was significantly associated with change in ODI at follow-up was type of treatment (non-operative or TDR) (B=7.2, 95% CI 0.5-13.8, p=0.04). Therefore, we analyzed each treatment group separately. However, we did not find any significant association between increased rating in any ADD variable and change in ODI (R^2=0.06 (p=0.85) in patients treated non-operatively and R^2=0.10 (p=0.50) in patients treated with TDR). Multiple logistic regression analysis did not reveal any association between the increase in any ADD variable and ODI ≤ 22 points.

The two treatment groups did not differ significantly in regards to the proportion of patients with increased rating value for one, two, three or four ADD variables (p=0.38) (Figure 3).

Discussion

This eight-year follow-up of the first randomized trial to compare TDR with non-operative treatment revealed no difference in ADD development between the treatment groups. This supports the theory that ADD development is part of the natural course of degeneration. Furthermore, the ADD development was not related to the clinical outcome.

In each treatment group, a larger proportion of patients had increased ADD at the eight-year follow-up than at the two-year follow-up (40% versus 19% after non-operative treatment and 42% versus 13% after TDR) [7].) A further increase is expected with longer follow-up.

Reduced range of motion (ROM) in the prosthesis has previously been reported to be associated with an increased prevalence of ADD [22 23]. We did not measure ROM at eight-year follow-up, but at two-year follow-up [24] segmental ROM was similar for an average disc prosthesis as for a degenerated index level disc. This may have contributed to similar ADD development in both treatment groups.

In a previous review of ADD following back surgery, Harrop et al [25] found a large variation in the reported ADD (0-24 % of the patients 3 to 17 years after TDR). However, the included studies were heterogeneous and had major limitations. The large variation in ADD most likely reflects differences in patient characteristics (e.g. age), follow-up time and ADD assessment methods. Such methods were either not reported or included disc height, osteophyte formation or instability on flexion-extension images. The variation in ADD assessment makes it difficult to compare the proportions of patients with ADD increase, also between recent studies. Zigler et al [26] based ADD on disc height reduction, endplate sclerosis, osteophytes and

spondylolisthesis on radiographs. They found increased ADD in 9% of patients five years after TDR. We found a much higher proportion (42%) with increased ADD at eight-year follow-up. The difference may partly be due to our use of MRI to detect changes not visible on radiographs (Table 3).

Regardless of allocation, 91 patients in our cohort have been treated with TDR [8], of whom one (1%) has been re-operated with fusion including the adjacent segment. This is in agreement with other studies reporting re-operation due to ADD in 0-1% of patients after two to five years following TDR [27-29].

We found no significant association between increased ADD and the clinical outcome at eight-year follow-up. This is in line with a previous report from Huang et al [22] and with the results of a recent cross-sectional analysis of long-term follow-up data from four randomized trials comparing non-operative treatment with fusion for chronic LBP [6].

The main strength of this study is its original randomized design that allows for prospective comparison of ADD development after TDR with the course of ADD in patients treated non-operatively. Other strengths are the public financing of the study, the long follow-up time, the metal artefact reducing MRI protocol, the MRI evaluation performed independently by experienced radiologists blinded to the clinical outcome and the direct comparison of post- and pre-treatment MRIs to assess changes in ADD. Such comparison can reduce overrating of changes due to ambiguous findings or small differences in MRI techniques and can improve agreement on changes in ratings [30]. In this study, the agreement was mostly good, despite instances where a low prevalence of change tended to reduce many of the kappa values. The conclusive combined ratings from both observers were likely to be even more reliable [31], reducing the chance of underestimating the MRI findings' relationship to other variables [32].

We studied a broad range of separate ADD variables. There are several ways to describe ADD, and no gold standard exists. The commonly used Pfirrmann system [33] provides a single rating of disc degeneration based on the height, structure and signal of the disc, and the distinction of nucleus and annulus. This system does not separate disc signal from disc height, and it does not include disc contour/herniation or HIZ – nor Modic changes, which were related to clinical outcome after TDR in our cohort in both the short- [34] and long-term [35]. Similarly, as at 2-year follow-up [7], we accepted an increase in only one ADD variable as indicating increased ADD. We were thus able to detect even small increases and differences in ADD.

The main limitations of the present study are its small sample size and that only 73% of the original sample could be included. This may call the generalizability of the results into question.

A further limitation is the choice of an as-treated analysis as the main analysis, since not all patients were analyzed according to the original randomization. However, the analyzed treatment groups had similar pre-treatment clinical, demographical and radiological characteristics. Since 24% of the patients randomized to rehabilitation were treated surgically, and 10% of those randomized to TDR were not operated [8], we considered an intention-to-treat analysis to be unsuitable for analyzing the influence of TDR on ADD.

The 11 patients who crossed over from rehabilitation to TDR had a shorter observation time. This may have reduced the proportion with increased ADD in the TDR group, but results were unchanged in the sensitivity analysis excluding these 11 patients.

The radiologists could not be blinded to the treatment group, which represents a possible observer bias. They were not blinded to post-treatment images when assessing pre-treatment images, and this may have influenced their pre-treatment ratings. Disc prostheses leave metal

artefacts on the images close to the implant, but, according to previous reports [36-38], such metal artefacts barely affect the evaluation of the adjacent disc level, and the metal artefact reducing MRI protocol further reduced the extent of the artefacts (Figure 2). The artefacts might hide the caudal part of large Modic changes extending caudally from the upper adjacent level, but we could still assess change in the size of Modic changes in all patients in our study.

The choice of the primary outcome variable may be debated. Increased ADD (yes/no) was a mainly qualitative variable. Only one of the six underlying ADD variables (disc height) was based on an actual measurement. Our study still provides more information than studies restricted to disc height measurement alone. By defining increase in ADD as a dichotomized variable, we did not use information on the degree of the increase. It is not clear how a variable reflecting the overall degree of increase in ADD can be constructed and weighted based on increases in different underlying variables (disc signal, disc contour, HIZ, etc.). However, we compared the number of increased ADD variables between groups in a post hoc analysis (Figure 3).

Several disc prostheses with different mechanical and geometrical properties are in use [39]. Different prosthesis designs allow different ROM, and the prosthesis used in this trial is classified as semi-constrained [39]. The mobility in the treated level can affect the development of ADD [22 23], and the development of ADD may have been different if another prosthesis design had been used.

In conclusion, the development of ADD at eight-year follow-up did not differ between non-operative treatment and TDR, and was not related to the clinical outcome. Hence, TDR as performed in the present study does not seem to increase the long-term risk of developing ADD, and patients developing ADD do not seem to have a worse outcome. Our results are similar with a recent study with long-term follow-up after lumbar fusion compared to non-

operative treatment [6], suggesting that other factors than fusion or TDR are important for the
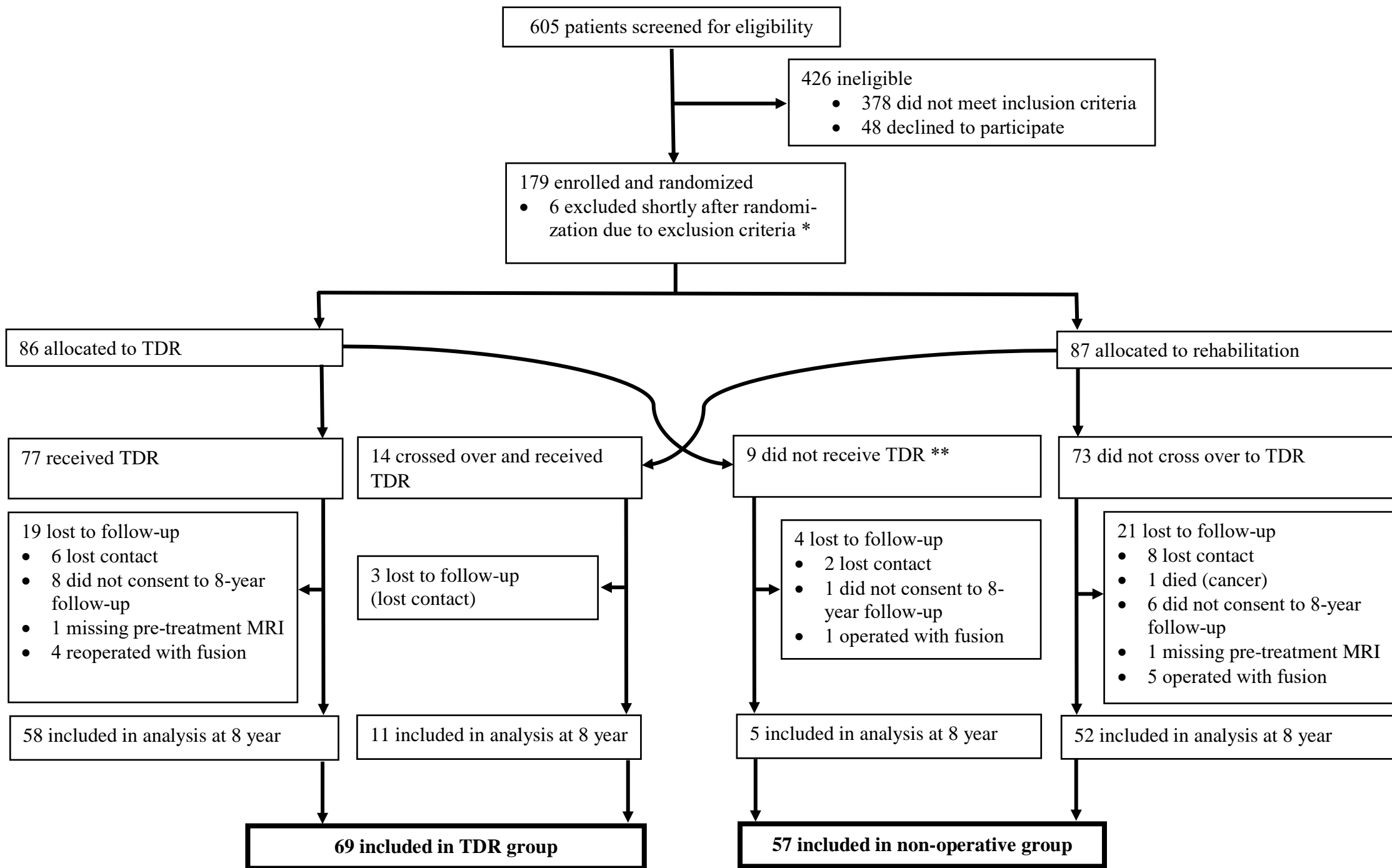
development of ADD.

References

1. Helgeson MD, Bevevino AJ, Hilibrand AS. Update on the evidence for adjacent segment degeneration and disease. The spine journal : official journal of the North American Spine Society 2013;**13**(3):342-51 doi: 10.1016/j.spinee.2012.12.009[published Online First: Epub Date]|.
2. Jacobs W, Van der Gaag NA, Tuschel A, et al. Total disc replacement for chronic back pain in the presence of disc degeneration. Cochrane database of systematic reviews (Online) 2012;**9**:CD008326 doi: 10.1002/14651858.CD008326.pub2[published Online First: Epub Date]|.
3. Omair A, Mannion AF, Holden M, et al. Age and pro-inflammatory gene polymorphisms influence adjacent segment disc degeneration more than fusion does in patients treated for chronic low back pain. Eur. Spine J. 2016;**25**(1):2-13 doi: 10.1007/s00586-015-4181-x[published Online First: Epub Date]|.
4. Song KJ, Choi BW, Jeon TS, Lee KB, Chang H. Adjacent segment degenerative disease: is it due to disease progression or a fusion-associated phenomenon? Comparison between segments adjacent to the fused and non-fused segments. Eur. Spine J. 2011;**20**(11):1940-5 doi: 10.1007/s00586-011-1864-9[published Online First: Epub Date]|.
5. Lund T, Oxland TR. Adjacent level disk disease--is it really a fusion disease? Orthop. Clin. North Am. 2011;**42**(4):529-41, viii doi: 10.1016/j.ocl.2011.07.006[published Online First: Epub Date]|.
6. Mannion AF, Leivseth G, Brox JI, Fritzell P, Hagg O, Fairbank JC. ISSLS Prize winner: Long-term follow-up suggests spinal fusion is associated with increased adjacent segment disc degeneration but without influence on clinical outcome: results of a combined follow-up from 4 randomized controlled trials. Spine 2014;**39**(17):1373-83 doi: 10.1097/brs.0000000000000437[published Online First: Epub Date]|.
7. Hellum C, Berg L, Gjertsen O, et al. Adjacent level degeneration and facet arthropathy after disc prosthesis surgery or rehabilitation in patients with chronic low back pain and degenerative disc: second report of a randomized study. Spine 2012;**37**(25):2063-73 doi: 10.1097/BRS.0b013e318263cc46[published Online First: Epub Date]|.
8. Furunes H, Storheim K, Brox JI, et al. Total disc replacement versus multidisciplinary rehabilitation in patients with chronic low back pain and degenerative discs: 8-year follow-up of a randomized controlled multicenter trial. The spine journal : official journal of the North American Spine Society 2017;**17**(10):1480-88 doi: 10.1016/j.spinee.2017.05.011[published Online First: Epub Date]|.
9. Hellum C, Johnsen LG, Storheim K, et al. Surgery with disc prosthesis versus rehabilitation in patients with low back pain and degenerative disc: two year follow-up of randomised study. BMJ 2011;**342**:d2786 doi: 10.1136/bmj.d2786[published Online First: Epub Date]|.
10. Brox JI, Sorensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. Spine 2003;**28**(17):1913-21 doi: 10.1097/01.brs.0000083234.62751.7a[published Online First: Epub Date]|.
11. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine 2000;**25**(22):2940-52; discussion 52
12. Grotle M, Brox JI, Vollestad NK. Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. J. Rehabil. Med. 2003;**35**(5):241-7

13. Marshman LA, Strong G, Trewhella M, Kasis A, Friesem T. Minimizing ferromagnetic artefact with metallic lumbar total disc arthroplasty devices at adjacent segments: technical note. Spine 2010;**35**(2):252-6 doi: 10.1097/BRS.0b013e3181c838cc[published Online First: Epub Date]|.

14. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 1988;**166**(1 Pt 1):193-9

15. Jensen TS, Sorensen JS, Kjaer P. Intra- and interobserver reproducibility of vertebral endplate signal (modic) changes in the lumbar spine: the Nordic Modic Consensus Group classification. Acta Radiol. 2007;**48**(7):748-54 doi: 10.1080/02841850701422112[published Online First: Epub Date]|.

16. Masharawi Y, Kjaer P, Bendix T, et al. The reproducibility of quantitative measurements in lumbar magnetic resonance imaging of children from the general population. Spine 2008;**33**(19):2094-100 doi: 10.1097/BRS.0b013e31817f19f7[published Online First: Epub Date]|.

17. 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 2014;**39**(24):E1448-65 doi: 10.1097/BRS.0b013e3182a8866d[published Online First: Epub Date]|.

18. Luoma K, Riihimaki H, Luukkonen R, Raininko R, Viikari-Juntura E, Lamminen A. Low back pain in relation to lumbar disc degeneration. Spine 2000;**25**(4):487-92

19. Aprill C, Bogduk N. High-intensity zone: a diagnostic sign of painful lumbar disc on magnetic resonance imaging. Br. J. Radiol. 1992;**65**(773):361-9 doi: 10.1259/0007-1285-65-773-361[published Online First: Epub Date]|.

20. Altman D. *Practical statistics for medical research*. New York: Chapman & Hall, 1991.

21. van Hooff ML, Mannion AF, Staub LP, Ostelo RW, Fairbank JC. Determination of the Oswestry Disability Index score equivalent to a "satisfactory symptom state" in patients undergoing surgery for degenerative disorders of the lumbar spine-a Spine Tango registry-based study. The spine journal : official journal of the North American Spine Society 2016;**16**(10):1221-30 doi: 10.1016/j.spinee.2016.06.010[published Online First: Epub Date]|.

22. Huang RC, Tropiano P, Marnay T, Girardi FP, Lim MR, Cammisa FP, Jr. Range of motion and adjacent level degeneration after lumbar total disc replacement. The spine journal : official journal of the North American Spine Society 2006;**6**(3):242-7 doi: 10.1016/j.spinee.2005.04.013[published Online First: Epub Date]|.

23. Putzier M, Funk JF, Schneider SV, et al. Charite total disc replacement--clinical and radiographical results after an average follow-up of 17 years. Eur. Spine J. 2006;**15**(2):183-95 doi: 10.1007/s00586-005-1022-3[published Online First: Epub Date]|.

24. Johnsen LG, Brinckmann P, Hellum C, Rossvoll I, Leivseth G. Segmental mobility, disc height and patient-reported outcomes after surgery for degenerative disc disease: a prospective randomised trial comparing disc replacement and multidisciplinary rehabilitation. The bone & joint journal 2013;**95-B**(1):81-9 doi: 10.1302/0301-620x.95b1.29829[published Online First: Epub Date]|.

25. Harrop JS, Youssef JA, Maltenfort M, et al. Lumbar adjacent segment degeneration and disease after arthrodesis and total disc arthroplasty. Spine 2008;**33**(15):1701-7 doi: 10.1097/BRS.0b013e31817bb956[published Online First: Epub Date]|.

26. Zigler JE, Glenn J, Delamarter RB. Five-year adjacent-level degenerative changes in patients with single-level disease treated using lumbar total disc replacement with ProDisc-L versus circumferential fusion. J. Neurosurg. Spine 2012;**17**(6):504-11 doi: 10.3171/2012.9.spine11717[published Online First: Epub Date]|.

27. Berg S, Tullberg T, Branth B, Olerud C, Tropp H. Total disc replacement compared to lumbar fusion: a randomised controlled trial with 2-year follow-up. Eur. Spine J. 2009;**18**(10):1512-9 doi: 10.1007/s00586-009-1047-0[published Online First: Epub Date]|.

28. McAfee PC, Geisler FH, Saiedy SS, et al. Revisability of the CHARITE artificial disc replacement: analysis of 688 patients enrolled in the U.S. IDE study of the CHARITE Artificial Disc. Spine 2006;**31**(11):1217-26 doi: 10.1097/01.brs.0000217689.08487.a8[published Online First: Epub Date]|.

29. Guyer RD, McAfee PC, Banco RJ, et al. Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: five-year follow-up. The spine journal : official journal of the North American Spine Society 2009;**9**(5):374-86 doi: 10.1016/j.spinee.2008.08.007[published Online First: Epub Date]|.

30. Berg L, Neckelmann G, Gjertsen O, et al. Reliability of MRI findings in candidates for lumbar disc prosthesis. Neuroradiology 2012;**54**(7):699-707 doi: 10.1007/s00234-011-0963-y[published Online First: Epub Date]|.

31. Espeland A, Vetti N, Krakenes J. Are two readers more reliable than one? A study of upper neck ligament scoring on magnetic resonance images. BMC Med. Imaging 2013;**13**:4 doi: 10.1186/1471-2342-13-4[published Online First: Epub Date]|.

32. Jarvik JG, Deyo RA. Moderate versus mediocre: the reliability of spine MR data interpretations. Radiology 2009;**250**(1):15-7 doi: 10.1148/radiol.2493081458[published Online First: Epub Date]|.

33. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine 2001;**26**(17):1873-8

34. Hellum C, Johnsen LG, Gjertsen O, et al. Predictors of outcome after surgery with disc prosthesis and rehabilitation in patients with chronic low back pain and degenerative disc: 2-year follow-up. Eur. Spine J. 2012;**21**(4):681-90 doi: 10.1007/s00586-011-2145-3[published Online First: Epub Date]|.

35. Furunes H, Hellum C, Brox JI, et al. Lumbar total disc replacement: predictors for long-term outcome. Eur. Spine J. 2017 doi: 10.1007/s00586-017-5375-1[published Online First: Epub Date]|.

36. Neal CJ, Rosner MK, Kuklo TR. Magnetic resonance imaging evaluation of adjacent segments after disc arthroplasty. J. Neurosurg. Spine 2005;**3**(5):342-7 doi: 10.3171/spi.2005.3.5.0342[published Online First: Epub Date]|.

37. Yang CW, Liu L, Wang J, et al. Magnetic resonance imaging of artificial lumbar disks: safety and metal artifacts. Chin. Med. J. (Engl.) 2009;**122**(8):911-6

38. Berg L, Gjertsen O, Hellum C, et al. Reliability of change in lumbar MRI findings over time in patients with and without disc prosthesis--comparing two different image evaluation methods. Skeletal Radiol. 2012;**41**(12):1547-57 doi: 10.1007/s00256-012-1394-5[published Online First: Epub Date]|.

39. Galbusera F, Bellini CM, Zweig T, et al. Design concepts in lumbar total disc arthroplasty. Eur. Spine J. 2008;**17**(12):1635-50 doi: 10.1007/s00586-008-0811-x[published Online First: Epub Date]|.

Figure 1: **Trial profile**

```
                         ┌─────────────────────────────────────┐
                         │  605 patients screened for eligibility │
                         └─────────────────────────────────────┘
                                        │
                                        │        ┌──────────────────────────────────────┐
                                        ├───────▶│  426 ineligible                        │
                                        │        │   • 378 did not meet inclusion criteria │
                                        │        │   • 48 declined to participate          │
                                        │        └──────────────────────────────────────┘
                                        ▼
                         ┌─────────────────────────────────────────┐
                         │  179 enrolled and randomized              │
                         │   • 6 excluded shortly after randomi-     │
                         │     zation due to exclusion criteria *    │
                         └─────────────────────────────────────────┘
```

| 86 allocated to TDR | | | 87 allocated to rehabilitation |

| 77 received TDR | 14 crossed over and received TDR | 9 did not receive TDR ** | 73 did not cross over to TDR |

| 19 lost to follow-up
• 6 lost contact
• 8 did not consent to 8-year follow-up
• 1 missing pre-treatment MRI
• 4 reoperated with fusion | 3 lost to follow-up (lost contact) | 4 lost to follow-up
• 2 lost contact
• 1 did not consent to 8-year follow-up
• 1 operated with fusion | 21 lost to follow-up
• 8 lost contact
• 1 died (cancer)
• 6 did not consent to 8-year follow-up
• 1 missing pre-treatment MRI
• 5 operated with fusion |

| 58 included in analysis at 8 year | 11 included in analysis at 8 year | 5 included in analysis at 8 year | 52 included in analysis at 8 year |

| **69 included in TDR group** | | **57 included in non-operative group** |

*Included and randomized mistakenly, but excluded before intervention because of obvious exclusion criteria

**Patients chose not to receive allocated treatment

Figure 2. Increased degeneration of the adjacent disc from baseline to eight-year follow-up after non-operative treatment (1-2) and total disc replacement (3-4).
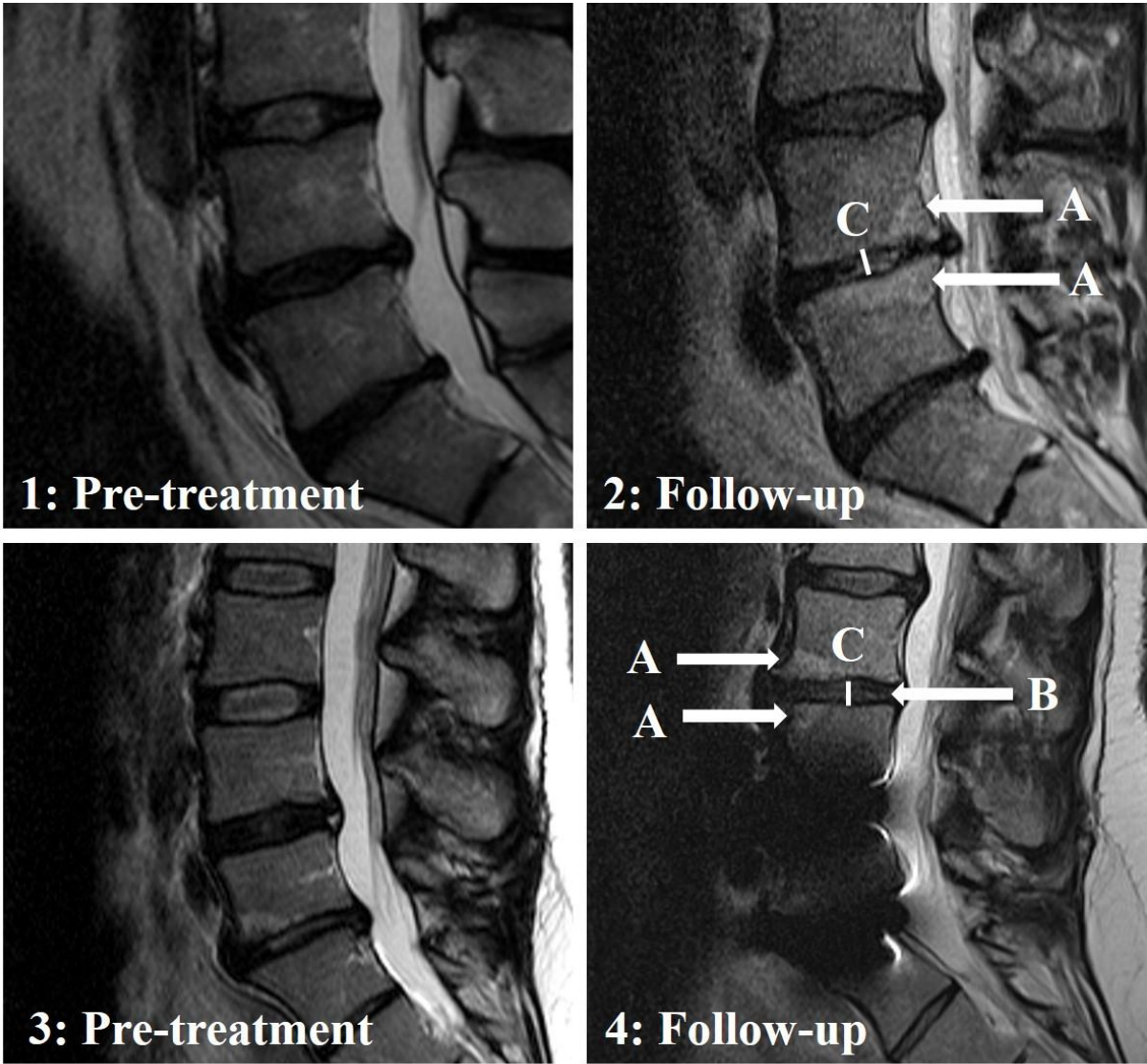
A=Modic changes, B=changed signal intensity of the disc and C=reduced disc height.

Figure 3. Proportions of patients with different numbers of variables showing increased adjacent disc degeneration.
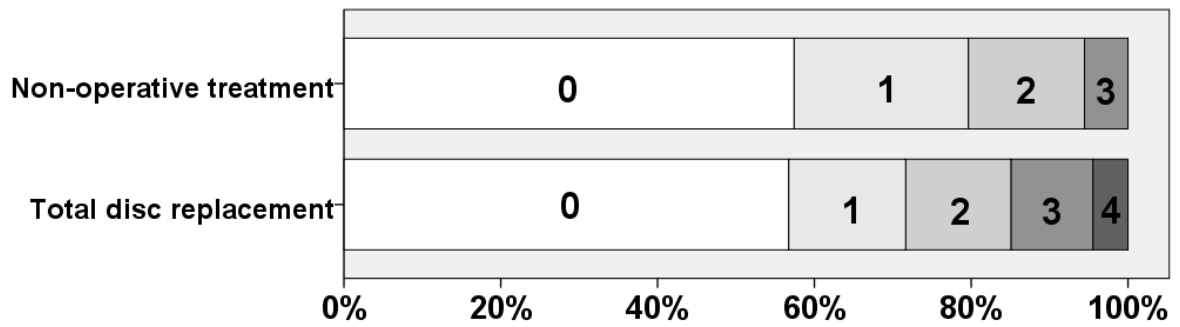
Proportions of patients treated non-operatively or with total disc replacement with increased degeneration in the adjacent disc in none, one, two, three or four out of six variables (Modic changes, disc height, disc contour, herniation size, nucleus pulposus signal and posterior high intensity zone) were not significantly different (p=0.38).

Table 1. The MRI variables that the evaluation of adjacent disc degeneration was based on.

Variable	Description	Grading	Reference
Type of Modic changes	Modic changes in the vertebral body marrow adjacent to the endplate.	0: No change Type 1: Hypointense T1-signal and hyperintense T2-signal Type 2: Hyperintense T1-signal and iso- or hyperintense T2-signal Type 3: Hypointense T1-signal and hypointense T2-signal	Modic 1998 (14)
Extent of Modic changes	The maximal craniocaudal extent of Modic changes. Vertebral body height was assessed on the image showing the largest height of Modic changes.	0: No signal changes 1: < ¼ of the vertebral body height 2: ¼-½ of the vertebral body height 3: > ½ of the vertebral body height	Jensen 2007 (15)
Disc height	The distance between the mid-inferior and the mid-superior disc borders, measured in millimetres on the mid-sagittal T2-weighted image.	0: No change 1: Disc height reduction ≥ 2 millimeters *	Masharawi 2008 (16)
Disc contour	The shape of the disc.	0: Normal 1: Bulging (> 1/4 of disc circumference) 2: Herniated (including protrusion, extrusion, and sequestration)	Fardon 2014 (17)
Size of herniation	Herniation size compared to the cross-sectional area of the spinal canal on axial view.	0: No herniation 1: <1/3 of cross-sectional area of the spinal canal 2: 1/3-2/3 of cross-sectional area of the spinal canal 3: > 2/3 of cross-sectional area of the spinal canal	Fardon 2014 (17)
Nucleus pulposus signal	Visually graded on sagittal T2-weighted images, using cerebrospinal fluid as intensity reference.	0: Bright 1: Grey 2: Dark 3: Black	Luoma 2000 (18)
Posterior high intensity zone (HIZ)	An area of high-signal intensity in the posterior annulus fibrosus, brighter than or equally as bright as cerebrospinal fluid on sagittal T2-weighted images, and surrounded superiorly, inferiorly and anteriorly by the low-intensity signal of the annulus fibrosus.	0: Not present 1: Present	Aprill 1992 (19)

* The smallest detectable change in disc height has previously been calculated as 2 mm (Hellum 2006 (7)). Thus, disc height reduction < 2 mm from pre-treatment to follow-up was rated as «no change».

Table 2: Pre-treatment clinical, demographical and radiological characteristics in patients treated non-operatively (n=57) or operated with total disc replacement (n=69).

	No prosthesis	Missing	Prosthesis	Missing	p
Clinical / demographical variables					
Age (median, range)	43 (29-54)	0	40 (25-54)	0	0.30
Gender (female) (n, %)	33 (58)	0	33 (48)	0	0.29
Comorbidity (n, %)	12 (21)	0	15 (22)	0	1.00
Body mass index (median, range)	25.7 (19.7-33.2)	0	25.0 (18.5-35.4)	1	0.30
Current smoker (n, %)	23 (40)	0	30 (44)	1	0.72
Previous surgery (n, %)	15 (26)	0	21 (30)	0	0.69
Index level	59	0	69	0	0.16
L4/L5 (n, %)	10 (18)		13 (19)		
L5/S1 (n, %)	37 (63)		33 (48)		
L4/L5 and L5/S1 (n, %)	11 (19)		23 (33)		
ODI* (median, range)	42 (30-66)	0	40 (28-68)	0	0.77
Back pain (VAS**) (median, range)	73 (50-98)	2	70 (19-97)	2	0.09
Radiological variables * **					
Pelvic incidence (median, range)	50 (33-71)	3	50 (25-75)	2	0.71
Modic changes	57	0	69	0	0.96
Not present (n, %)	53 (93)		65 (94)		
Type 1 (n, %)	1 (2)		1 (1)		
Type 2 (n, %)	3 (5)		3 (4)		
Adjacent disc height (mm) (median, range)	10 (6-14)	0	10 (8-14)	0	0.63
Disc contour	57	0	69	0	0.16
Normal (n, %)	39 (68)		57 (83)		
Bulge (n, %)	12 (21)		9 (13)		
Herniation (n, %)	6 (11)		3 (4)		
Nucleus pulposus signal §	57	0	69	0	0.58
Bright (n, %)	25 (44)		35 (51)		
Grey (n, %)	20 (35)		24 (35)		
Dark (n, %)	12 (21)		10 (14)		
Posterior high intensity zone (n, %)	3 (5)	0	4 (6)	0	0.90

* ODI=Oswestry Disability Index (0 to 100, lower scores indicate less severe symptoms)

** VAS=Visual Analogue Scale (Calculated using a horizontal scale ranging from 0 (no pain) to 100 (worst pain imaginable), with word anchors at the beginning and end).

*** All radiological variables are measured by MRI at the upper adjacent level, except pelvic incidence, which is measured on plain radiograph with a lateral view of the patient in standing position.

§ Luoma et al. (18)

Table 3: Radiological measures at adjacent discs L3/L4 or L4/L5 before treatment and at eight-year follow-up in 57 patients treated non-operatively and 69 patients with total disc replacement.

	Non-operative treatment			Total disc replacement			P *
	Pre-treatment	Long-term follow-up	Change	Pre-treatment	Long-term follow-up	Change	
Modic changes	57	57		69	69		0.98
Not present (n, %)	53 (93)	48 (84)	-5	65 (94)	59 (86)	-6	
Type 1 (n, %)	1 (2)	1 (2)	0	1 (1)	6 (9)	+5	
Type 2 (n, %)	3 (5)	7 (12)	+4	3 (4)	4 (6)	+1	
Both type 1 and 2 (n, %)	0	1 (2)	+1	0	0	0	
Maximal craniocaudal extension of Modic changes (n, %)	57	57		69	69		0.94
No Modic changes	53 (93)	48 (84)	-5	65 (94)	60 (8)	-5	
< ¼ of vertebral body height	1 (2)	4 (7)	+3	2 (3)	1 (1)	-1	
¼ - ½ of vertebral body height	2 (4)	4 (7)	+2	1 (1)	4 (6)	+3	
> ½ of vertebral body height	1 (1)	1(1)	0	1 (1)	4 (6)	+3	
Adjacent disc height (mm) (mean, 95 % CI)	10.1 (9.6-10.5)	9.9 (9.3-10.4)	0.2 (0.0-0.4)	10.2 (9.9-10.5)	9.7 (9.3-10.1)	0.5 (0.2-0.7)	0.052
Adjacent disc height reduction ≥ 2 mm from pretreatment to long-term follow-up (n, %)		3 (5)			11 (16)		0.09
Disc contour	57	57		69	69		0.53
Normal (n, %)	39 (68)	28 (49)	-11	57 (83)	41 (59)	-16	
Bulge (n, %)	12 (21)	18 (32)	+6	9 (13)	20 (29)	+11	
Herniation (n, %)	6 (11)	11 (19)	+5	3 (4)	8 (12)	+5	
Herniation size	57	57		69	69		0.38
< 1/3 of spinal canal	6 (11)	10 (18)	+4	3 (4)	6 (9)	+3	
1/3-2/3 of spinal canal	0	1 (2)	+1	0	2 (3)	+2	
> 2/3 of spinal canal	0	0	0	0	0	0	
Nucleus pulposus signal **	57	57		69	69		0.68
Bright (n, %)	25 (44)	20 (35)	-5	35 (51)	22 (32)	-13	
Grey (n, %)	20 (35)	18 (32)	-2	24 (35)	30 (43)	+6	
Dark (n, %)	12 (21)	19 (33)	+7	10 (14)	17 (25)	+7	
Posterior high intensity zone (n, %)	3 (5)	5 (9)	+2	4 (6)	5 (7)	+1	0.45

* p values (Pearson Chi-Square test for differences in changes for the worse when comparing the two groups)

** Luoma et al. (18)