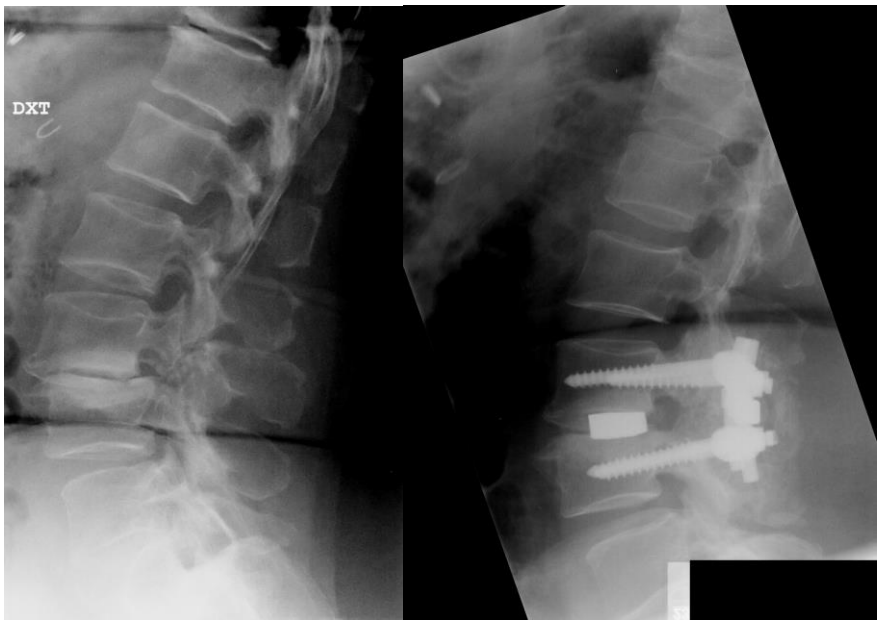
AARHUS UNIVERSITY

TLIF versus PLF in Degenerative Lumbar Spinal Disorders - Short- and Long-term PROMS, Pain, and Health Economics - Results from an RCT

PhD dissertation

Kristian Høy



Health

Aarhus University

2017

TLIF versus PLF in Degenerative Lumbar Spinal Disorders - Short- and Long-term PROMS, Pain, and Health Economics - Results from an RCT

PhD dissertation

Kristian Høy

Faculty of Health Science

Department of Orthopedics E, Spine Unit, AUH

Aarhus University

Correspondence

Kristian Høy, MD

Consultant, associate professor

Department of Orthopedics, Spine Unit.

Aarhus University Hospital

Nørrebrogade 44, 8000 Aarhus C

Phone: +45 20105058

E mail: krishoey@rm.dk, wintherhoey@dadlnet.dk

Table of contents

Preface and acknowledgements

This thesis is based on research done during my appointment as chief surgeon at the Spine Unit, Department of Orthopedics, Aarhus University Hospital, Denmark.

The Ph.D project was supervised by

Professor Cody Eric Bünger, MD, DMSc.

Associate professor Thomas Borbjerg Andersen, MD, PhD, DMSc

Professor Rikke Søgaard, MPh, MSc, PhD

My acknowledgements go to Professor Cody Bünger, who for many years taught me about - and included me in orthopedic research continuously ongoing in our department. The project was initiated in 2003, and I was appointed as the main researcher alongside my daily clinical work.

The research was possible due to the effort made by my colleagues, the surgeons at the Spine Unit Aarhus University Hospital: Bent Niedermann MD, Peter Helmig MD, PhD, Ebbe Stender Hansen MD, DMSc, Haisheng Li, MD, PhD, who all contributed to the project by inclusion and operation of the patients in order to fulfill the calculated sample size during the inclusion period from November 2003 until November 2008. Without their effort, this thesis would not have been possible.

Casper Saebye MD, PhD student, who always was helpful, when my computer wouldn't cooperate.

Special thanks go to Thomas Borbjerg Andersen, with whom I spent several hours in the operating theater watching him grow from a medical student to a full-grown candidate for a professorship in orthopedic spine surgery, for helping me to keep track of all the data.

My thanks go to Rikke Søgaard for bringing in a new perspective into daily clinical practice and trying to teach us clinicians how to spend money in a reasonable way for the community and all the taxpayers and, in the end, to optimize the health care system.

Frank Farsø Nielsen MD, PhD, the former head of the Orthopedic Department E, is acknowledged for providing me with a weekly day off from normal clinical duty for a period of one year in order to manage and interpret the huge amount of data.

My colleague and coworker for almost 30 years Ebbe Stender Hansen did a great and remarkable work in tuning the final manuscript with suggestions and linguistics.

I am thankful to Clinical Professor and Chair of Orthopedics at Aarhus University, Kjeld Søballe MD, DMSc, who made it possible for me to submit this thesis for evaluation at the Faculty of Health Sciences, Aarhus University.

Mette, my dearest soulmate and wife, provided my mind with the peace, which made this thesis possible.

List of papers

This thesis is based on four original papers.

Study 1

Kristian Høy, Cody Bünger, Bent Niedermann, Peter Helmig, Ebbe Stender Hansen, Haisheng Li, Thomas Andersen. Transforaminal lumbar interbody fusion (TLIF) versus posterolateral instrumented fusion (PLF) in degenerative lumbar disorders: a randomized clinical trial with 2-year follow-up. Eur Spine J. 2013 Sep; 22(9):2022-9. doi: 10.1007/s00586-013-2760-2. Epub 2013 Apr 13.PMID:23584162 **Free PMC Article**

Study 2

Kristian Høy, Kamilla Troung, Thomas Andersen, Cody Bünger. Addition of TLIF does not improve outcome over standard posterior instrumented fusion. 5-10 years long-term Follow-up: results from a RCT. Eur Spine J. 2017 Mar; 26(3):658-665. doi: 10.1007/s00586-016-4592-3. Epub 2016 May 7. PMID: 27155825

Study 3

Kristian Høy, Blazej Grycel, Thomas Andersen, Cody Bünger. Residual radiculopathy in TLIF surgery is not related to the side of cage insertion, but is higher in comparison to an instrumented standard posterolateral fusion measured by pain drawings. Results from a RCT. Submitted.

Study 4

Ann Dement Christensen, Kristian Høy, Cody Bünger, Peter Helmig, Ebbe Stender Hansen, Thomas Andersen, Rikke Søgaard. Transforaminal lumbar interbody fusion vs. posterolateral instrumented fusion - Cost-utility evaluation along side an RCT with 2-years of follow-up. Eur Spine J. 2014 May;23(5):1137-43. doi: 10.1007/s00586-014-3238-6. Epub 2014 Feb 21.

PMID:24557326

The papers will be referred by their numbers (Study 1, Study 2, Study 3, Study 4).

The author of this thesis presented all the papers orally, at the following national and international meetings:

Study 1:

1) 55th Nordic Orthopaedic Federation (NOF) May 2010. *Selected for the contest "Best Paper Award".*
2) The Seventh SICOT/SIROT Annual International Conference, a combined meeting with the Swedish Orthopaedic Association Gothenburg, Sweden, September 2010.
3) Danish Orthopaedic Association (DOS), Copenhagen, Denmark. October 2010. *Selected for participation in "Best Paper Award Contest".*
4) Danish Orthopaedic Association (DOS), Copenhagen, Denmark. October 2011.

5) The third Spine Week, Eurospine, Amsterdam, the Netherlands, May 2012.
6) Combined 33rd SICOT & 17th PAOA Orthopaedic World Conference Dubai, United Arab Emirates, November 2012.

Study 2:

1) Eurospine 2015, Selected as participant in "*Best Paper Award Contest*".
2) Danish Orthopedic Association (DOS), Copenhagen, October 2015.

Study 3:

1) Eurospine, Liverpool, October 2013
2) Danish Orthopedic Association (DOS), October 2013.
3) 44th Annual Meeting ISSLS, Seoul, June 2014

Study 4:

1) 44th Annual Meeting ISSLS, Seoul, June 2014

Abstract

This PhD dissertation seeks to explore the surgical technique Transforaminal Interbody Lumbar Fusion (TLIF) as a new treatment method for a population of patients with chronic low back pain (CLBP) due to degenerative lumbar disorders in a short- and long-term perspective.

The dissertation contains short-term (1 and 2 years) as well as long-term (average 9 years) results from a prospective randomized study conducted in a patient population selected for surgical treatment because of CLBP due to degenerative lumbar disorders.

Patient-related outcome measurement in the group of patients operated on with TLIF are compared to that of "Instrumented posterolateral fusion" (PLF), which is commonly regarded as the gold standard for these patients.

The dissertation focuses on possible differences in patient-experienced treatment effects between the two methods. Recognized standardized, validated, and generic indicators of health status in form of patient-reported functional outcome scores are employed.

Additionally, an attempt is made to detect whether the TLIF technique poses a risk of permanent affection of the nerve structures exposed in connection with the surgery.

Finally, the dissertation examines the health economics of TLIF in a classical cost-utility evaluation.

It is investigated whether TLIF provides improved health per used resource unit in comparison with PLF. Thus, it is assessed, whether the method can be justified based on a socioeconomic consideration of ¨value for money¨.

The dissertation could not demonstrate a positive effect of the TLIF in patient-reported outcome measures over the current standard treatment, neither in the short- nor the long-term.

The TLIF operation method resulted in prolonged surgery time and increased perioperative bleeding.

The study showed that both treatments gave an improved quality of life in the study population of patients with degenerative lumbar disorders elected for surgical treatment. The improved quality of life achieved by the surgical intervention was still present 9 years after the index operation. The positive effect of surgery was less in patients who underwent additional surgery after the index operation.

An increased incidence of new or permanent pain related to insertion of the cage or the special approach in the group of patients exposed to TLIF could not be detected.

There was no evidence that TLIF improved health or lowered costs and thus increased health effect per used resource unit over that of PLF. Thus, the method was not cost-effective compared to a standard instrumented posterolateral fusion.

Abstract in Danish

Denne PhD afhandling undersøger den ”nye” kirurgiske teknik stivgørende Transforaminal interkorporal lænderygskirurgi (TLIF) i en population af patienter med kroniske rygsmerter på grund af degenerativ lænderygsygdom i et kort- og langtids perspektiv. Afhandlingen tager udgangspunkt i et kontrolleret behandlingsforsøg, hvor sammenligningen er foretaget med udgangspunkt i den hidtidige accepterede behandling instrumenteret posterolateral stivgørende lænderygsoperation (PLF).

Afhandlingen indeholder korttids-(1 og 2 år) samt langtids-(9 år) resultater fra en prospektiv randomiseret kontrolleret undersøgelse af en studiepopulation med degenerativ lænderygssygdom visiteret til kirurgisk behandling.

Patient-rapporterede målbare effekter hos gruppen af patienter opereret med TLIF sammenlignes med samme effektmål i forhold til den alment anerkendte standardbehandling ”instrumenteret stivgørende lænderygs operation” (PLF) som kirurgisk metode for patienter med kroniske lænderygsmerter.

Afhandlingen fokuserer på forskelle i patienternes egne vurderinger af den opnåede behandlingseffekt.

Til måling af effekten anvendes anerkendte standardiserede, valide, og generelle måleredskaber for smerte, livskvalitet og helbredstatus relevant for en population af patienter med degenerativ lænderygsygdom.

Samtidig forsøges afdækket, om teknikken grundet den anderledes anatomiske adgang udgør en potentiel risiko for nyopstået og varig skade på nervestrukturer, som frilægges i forbindelse med interventionen.

Slutteligt undersøger afhandlingen de sundhedsøkonomiske problemstillinger ved stivgørende TLIF som ny kirurgisk metode. Specifikt undersøges det om TLIF er en omkostningseffektiv kirurgisk teknik set ud fra et sundhedssektorperspektiv baseret på en sundhedsøkonomisk analyse af den randomiserede 2 års opfølgning. Der suppleres med kvalitetsjusterede leveår (QALY) som effektmål og registerdata fra de danske nationale sundhedsregistre med henblik på at estimere de samlede sundhedsomkostninger.

Afhandlingen kunne ikke påvise en positiv effekt af TLIF sammenlignet med PLF hverken i kort- eller langtids resultater.

Operationsmetoden resulterede i længere operationslængde og øget blødning.

Resultaterne i undersøgelsen viste, at begge behandlinger gav en forbedret livskvalitet i effektmål hos populationen af patienter med kroniske lænderygsmerter grundet degenerativ

lænderygsygdom visiteret til kirurgisk behandling. Den forbedrede livskvalitet opnået ved den kirurgiske intervention var fortsat tilstede 9 år efter det operative indgreb.

Effekten af kirurgi var mindre ved patienter som i forløbet gennemgik flere operationer.

Der kunne ikke påvises øget forekomst af nyopståede eller varige nervesmerter relateret til den specielle anatomiske adgang i gruppen af patienter udsat for TLIF.

TLIF viste sig ikke omkostningseffektiv i et dansk sundhedssektorperspektiv fordi teknikken medførte en tendens til øgede omkostninger uden nogen samtidig gevinst i form af forbedret livskvalitet.

Abbreviations

ALIF Anterior Lumbar Interbody Fusion

ASD Adjacent Segment Degeneration

CLBP Chronic Low Back Pain

CONSORT Consolidated Standards of Reporting Trails

DDD Degenerative Disc Disease

DPQ Dallas Pain Questionnaire

DRG Diagnosis Related Grouping

HMS Health Measurements Scales

ICER Incremental Cost-Effectiveness Ratio

LBPRS Low Back Pain Rating Scale and Index

LTFU Long Term Follow Up

MCID Minimal Clinical Important Difference

MCS Mental Health Component Summary

MRI Magnetic Resonance Imaging

ODI Oswestry Disability Index

NPLF Non-Instrumented Posterolateral Lumbar Fusion

NRS Back Pain Numeric Rating Scale

PASS Patient Acceptable Symptom State

PCS Physical Health Component Summary

PLF Instrumented Posterolateral Fusion

PLIF Posterior Lumbar Interbody Fusion

PSF Posterior Spinal Fusion

QALY Quality-Adjusted-Life-Year

RCT Randomized Clinical Trial

SF-36 Short Form 36 General Health Instrument

TLIF Transforaminal Interbody Fusion

TSRH® Pedicle Instrumentation from Medtronic

VAS Visual Analogue Scale

XLIF Extreme Lateral Interbody Fusion

Introduction

The pathophysiology of low back pain (LBP) is complex. It is usually defined as pain, muscle tension, or stiffness localized below the costal margin and above the inferior gluteal crease. It can be present with or without radicular pain [3]. The first can be a result of mechanical compression of nerve roots and dura. In some cases, both types of pain are present.

However, only a small portion of patients with long lasting LBP should end up with a spinal fusion and never before various non-surgical conservative procedures such as physiotherapy, exercise, stretching, and pharmacological treatment have been tried for a longer period [3, 4]. One must keep in mind, that the treatment selected always should reflect the underlying pathology and that all treatment is individual and cannot be standardized.

As much as 12-35% of the background population will each year experience LBP. In fact the majority, i.e. 70-85%, will at some time in life experience back pain [3, 5]. A distinction between low back pain (LBP) and chronic low back pain (CLBP) is mandatory in order to reduce the cost of back pain, which alone in Denmark is estimated to be more than 17 billion DKK annually [6, 7] and has a high impact on the individual patient [8]. The European Guidelines and the Danish National Board of Health define CLBP as back pain lasting for more than 3 months. Patients suffering from CLBP have severe difficulties returning to work after sick leave, and after only 6 months the probability of returning is less than 50 %, and after 2

years it appears to be lost [5]. Mostly, patients with long lasting CLBP are thought to be candidates for fusion procedures.

Spinal fusion has been around for at least 100 years, and it has evolved over time. Primarily it was used to treat patients suffering from Potts´s disease, later for scoliosis and fractures.

Today CLBP due to degenerative lumbar disorders, i.e. spinal stenosis, spondylolisthesis, failed back surgery syndrome, and degenerative disc disease (DDD), accounts for 75% of all spinal procedures [9]. The fast growing development in hardware and technologies has led to an increase in lumbar fusion procedures worldwide [10]. The etiology of CLBP is complex and is

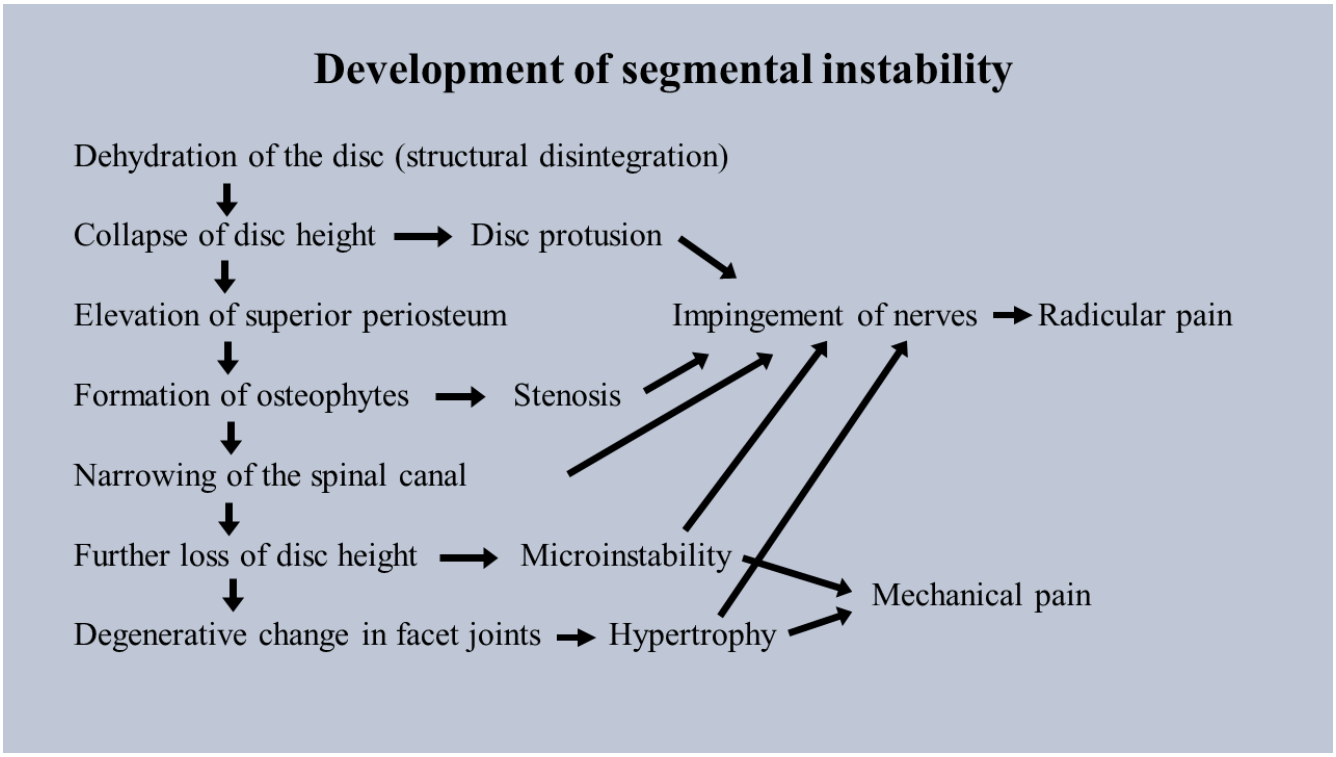
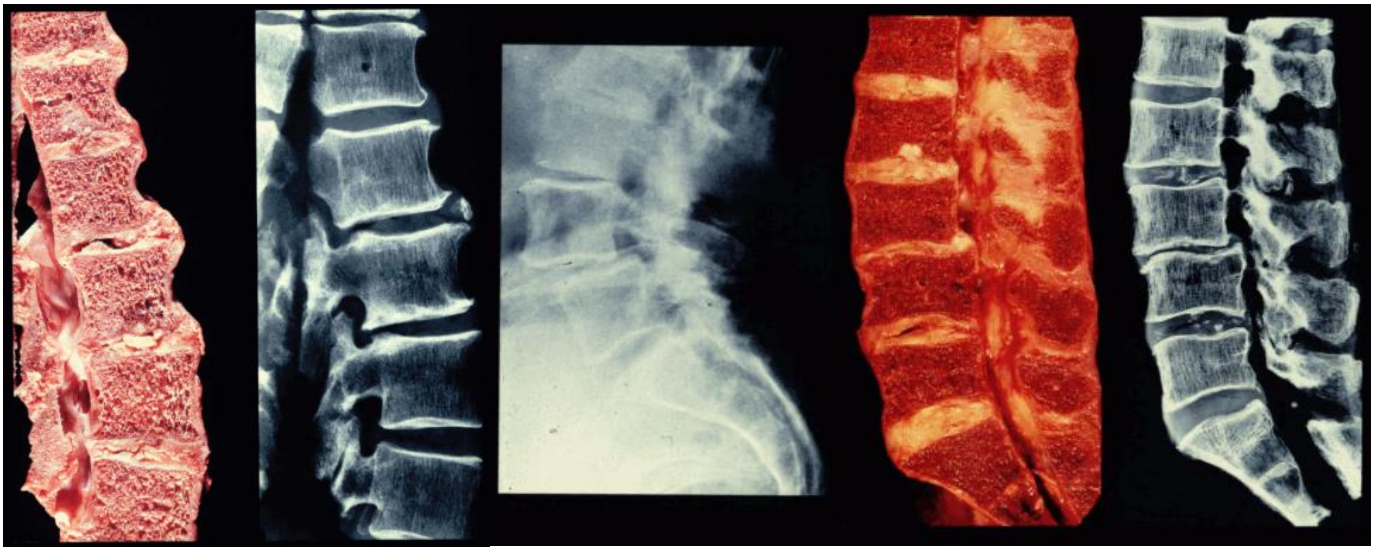


Figure 1. Pathology and pathogenesis of lumbar spondylosis modified by the author of this thesis after Kirkalaldy-Willis[2]

still not fully understood [11]. The pain generator can be the underlying disc degeneration caused by either normal ageing or genetic inheritance, which results in tissue weakening, tearing, and scarring of the annulus. Then a structural failure of the disc results in lowering of the disc height, and due to that the endplates of the vertebrae will come closer.



Picture 1. Anterior and posterior spondylolisthesis and osteophyte formation

As a result the pressure on facet joints increases, and facet joint osteoarthritis and hypertrophy ensues. The change in discs, facet joints, and ligaments can result in unstable spinal segments. Disc protrusion and hypertrophic facet joints can cause nerve impingement or entrapment and thereby pain due to spinal or foraminal stenosis and degenerative spondylolisthesis (Figure 1 and Picture 1).

The isthmic spondylolisthesis has its own nature. It is due to a defect in the pars interarticularis, which over time results in disc degeneration due to the sliding of the

vertebrae and stress on the adjacent disc resulting in fast and accelerated disc degeneration. The combination of reduced disc height and slippage results in nerve root entrapment in the portion of the nerve root canal immediately distal to the pedicle causing radiating pain.

Lumbar spinal fusion is often done where instability is present due to inheritance, previous decompression, or disc degeneration. The instability is thought to be a pain generator [12]. The rationale behind fusion is to stop motion in pathological motion segments and thereby to eliminate signals from facet joints, discs, and impinged nerves [13-16]. The literature on lumbar fusion consistently supports surgical treatment over conservative treatment of chronic low back pain [17-21]. Along the evolutionary trail and with gains of technological improvement different methods for achieving fusion have arisen. Various spinal techniques include three main types of fusion: 1) non-instrumented posterolateral fusion (NPLF), 2) Instrumented posterolateral lumbar fusion (PLF), and 3) Interbody lumbar fusion. The latter two methods are often combined [22-24].

Despite the high impact to society of CLBP, with a socioeconomic burden comparable to that of depression, heart disease, and diabetes[25], there still is no clear scientific evidence to refer to in selection of surgical strategy or even treatment strategy in general [26].

RCTs comparing instrumented and non-instrumented fusion have not been able to show any significant difference between the two surgical methods [19, 20, 27]. However, a newer study

showed significant improvement of an instrumented interbody fusion in comparison to a non-instrumented fusion [28].

Treatment strategies have over time moved toward global fusion. The movement toward interbody fusion is based on theoretical grounds, and the different techniques used in order to achieve global fusion is Posterior Lumbar Interbody Fusion (PLIF), Anterior Lumbar Interbody Fusion (ALIF), Transforaminal Interbody Lumbar Fusion (TLIF), and Extreme Lateral Interbody Fusion (XLIF).

The theoretical base of interbody fusion rests on the notion that restoration of disc height could indirectly decompress the neuroforamina, restore lost sagittal balance, and therefore in the long run result in improved outcome. The potential for a solid fusion is also thought to be higher since the fusion occurs over the weight-bearing axis of the spine. Addition of interbody fusion is recommended to enhance bony fusion [29].

In the PLIF procedure the disc is accessed by a posterior approach, which was first described by Briggs in 1944 [30]. The technique was in 1953 improved by Cloward using allograft dowels from the iliac crest [31] resulting in a higher spread of the method.

Nowadays interbody cages are mostly used [32]. However, the PLIF technique has some disadvantages in comparison to other interbody fusion techniques due to the canal violation and the need for dura- and nerve root retraction, which might lead to dural tear, canal fibrosis, and chronic radiculitis [33].

ALIF was also popularized by Cloward in the 50ies and had its high popularity in the nineties and early zeroes [19, 22, 32, 34]. The disc is accessed by a retroperitoneal approach and is often done in combination with posterior pedicle screws. However, the method is techniqually demanding due to operation close to the great vessel and prolonged operation time. Reports of retrograde ejaculation as high as 20 % [35, 36], and reports of high complication rates are matters of concern and interpretation [37].

In 1998, Harms described the TLIF method, where the disc was reached via a transforaminal route. The cage could be inserted without exposing more than the ipsilateral foramen. The entire facet was removed, which minimized the retraction of the dural sac, and the risk of fibrosis and scarring of neural structures could be reduced [38]. By placing the cage in the anterior or the middle part of the endplate and with pedicle screw instrumentation it could to a certain extent be possible to restore the lordosis over the segment. The method has won great popularity since it offered the theoretical benefits of 360 degrees fusion but with less morbidity due to a lesser surgical procedure, and it is nowadays used widely also as a minimally invasive TLIF. XLIF is also achieved by a minimal and extreme lateral approach.

The one and only systematic Cochrane review of lumbar fusion dates back to 2005 [26] and had an overall inclusion of 31 randomized clinical trials. In this review, eight RCTs showed an increased fusion in favor of instrumentation. However, the change in patient related outcome was marginal. This conclusion is close to the meta-analysis by Babu, who recently found that

there is evidence to suggest that lumbar instrumentation improves fusion rates, but no clear and consistent correlation has been found between fusion rate and patient related outcome [39]. In one of the latest guidelines (2014) for the performance of fusion procedures in the lumbar spine, the conclusion is *"that no general recommendations can … be made regarding the technique that should be used to achieve interbody fusion*" [29].

All though the long history and with the many different methods available, controversies on who, when, and how one should operate prevail. The lack of uniform treatment strategy is to a certain degree due to inconsistent interpretation of the diagnosis, symptoms, MRIs, and maybe skills, tradition, and culture in the institutions actively involved in the treatment.

The evidence for surgery compared to conservative treatment is scarce albeit present [19-21]. The long-term results comparing fusion to non-operative care has called a debate in the spine community. In 2013, Mannion concluded that fusion was not superior to conservative treatment [40]. Her conclusions derived from long term follow up of the two studies comparing fusion to conservative treatment originally initiated by Brox [18] and Fairbank [41]. However, Hedlund added to the controversies in 2016 when he concluded, that from the patients perspective lumbar fusion surgery is a valid treatment option for CLBP. The statement was built upon 12.8 years long term follow up of the Swedish lumbar spine study, which was initiated by Fritzell and coworkers in the late 90ies [19]. He changed the primary RCT into an observational study due to the many crossovers from the conservatively treated

group. In the RCT patients were scored after the “intention to treat principle” even though they were cross-overs and had a fusion procedure [42].

Despite the rising numbers of spinal fusions performed worldwide, the evidence for effect or value for money is not convincing concerning the different techniques used [37, 40, 42-47]. TLIF has become the most widely used interbody fusion method. However, the cost-effectiveness of this treatment modality is not known in a broader perspective and validated, and before it becomes the “gold standard in spinal fusion” it must prove its superiority in a clinical – as well as an economical perspective over the gold standard of today, which is a PLF.

This study was initiated in 2003. It was the ambition to provide the first comparison of an open TLIF to a standard PLF in a randomized prospective trial in order to provide Level I and II data to the spine communities concerning surgical treatment of CLBP, both regarding standard short term follow, long-term follow up (LTFU), and possible unexpected side effects with relation to the new method.

Hypothesis and aims

The aim of the present dissertation was to investigate TLIF as a new and rapidly growing fusion method compared to a standard instrumented posterior lumbar fusion (PLF) with special focus on functional outcome in a standard 2 years follow up and long-term (LTFU) perspective in patients suffering from (CLBP) pain due to degenerative lumbar disorders. Special focus was made to pain, pain localization, and level of pain. In addition, it was evaluated weather TLIF is cost-effective compared to PLF.

Study 1

Short-term standard Patient Reported Outcome Measurements (PROMS) in the TLIF cohort was compared to those of PLF in patients suffering from CLBP due to degenerative lumbar disorder.

H_0 is always that there is no difference between the groups.

H_1: It was hypothesized that the TLIF group would be superior in functional outcome parameters due to better stabilization, higher fusion rate, and indirect decompression of neuroforamina compared to the PLF group. So in order to accept H_1, H_0 must be rejected.

Study 2

Long-term PROMS in the TLIF group was compared to PLF in patients suffering from CLBP due to degenerative lumbar disorder.

H_0 is always that there is no difference between the groups.

H_1: It was hypothesized that the TLIF group over time would be superior in functional outcome measurements due to a higher fusion rate, better sagittal balance, restoration of lumbar lordosis and less reoperations since the method theoretically could restore disc height in the segments involved in the fusion and thereby reduce adjacent segment degeneration (ASD). Such effects have been seen in long-term results from other prospective randomized lumbar interbody fusion studies (ALIF) [47]. So in order to accept H_1, H_0 must be rejected.

Study 3

The study was undertaken to detect whether the TLIF procedure could result in an undesired side effect due to violation of the upper nerve root or ganglion resulting in chronic radiculopathy or radiculitis in the TLIF group compared to the PLF. In the TLIF group only it was also investigated whether there was a greater amount of radiculopathy at the side of cage insertion. With the aid of pain drawings it was investigated whether the TLIF method

resulted in newly developed radiculitis or radiculopathy compared to the PLF with no exposure of the nerve by pain drawings.

H_0 is always that there is no difference between the groups.

$H_{1:}$ It was hypothesized that, the TLIF group could develop more pain due to the special approach and especially the side of cage insertion could have a greater appearance of new neurogenic pain.

Study 4

To assess whether the TLIF was cost-effective in comparison to the gold standard of today, a PLF, with respect to QALYs gained compared to the expenses used.

H_0 is always that there is no difference between the groups.

H_1: It was hypothesized that the TLIF might be more cost-effective in terms of QALYs due to better outcome scores in the TLIF group. So in order to accept H_1, H_0 should be rejected.

Material and methods

From November 2003 through November 2008 a total of 100 patients (mean age 49.8 years, range 25-70 years; TLIF-group mean 50.3 years, PLF-group mean 49.3 years) were included in a single-center prospective randomized clinical trial.

All patients suffered from CLBP due to degenerative lumbar disorder with or without leg pain.

Inclusion criteria:

- Age $\geq$ 25 years
- Surgery at 1, 2 or 3 levels between L2 and S1
- Segmental instability or spinal stenosis and need for central or foraminal decompression
- Instability due to disc degeneration or secondary to earlier decompression or discectomy.
- Spondylolisthesis Meyerding grade 1 or 2

Exclusion criteria:

- Spondylolisthesis Meyerding grade 3 or 4
- Osteoporosis

- Significant cerebral or cardiovascular history
- Kidney problems
- Malignancy
- Use of drugs reducing bone metabolism
- Dementia, psychological abnormality
- Language problems

All patients were recruited in the outpatient clinic by senior surgeons.

Sample size

Sample size of the study was calculated using a significance level of 0.05 and a power of 0.80.

Based on earlier studies, the standard deviation (SD) of the DPQ daily activity score was set at 25 points. A 15-point difference in this category was considered as the minimally clinically important difference (MCID). To fulfill these criteria, the study would need 44 patients in each group. In order to allow for dropouts, the study was designed with 50 patients in each group. The equation used for power calculation was [48].

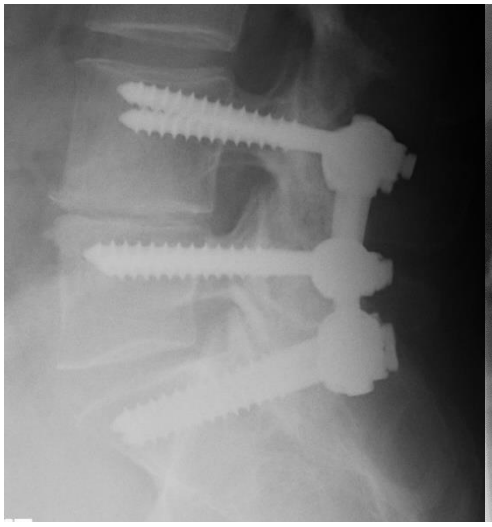
$$n_{\text{each group}} = 2\left((1{,}96+0{,}84)*25/15\right)^2 = 43{,}56$$

Randomization

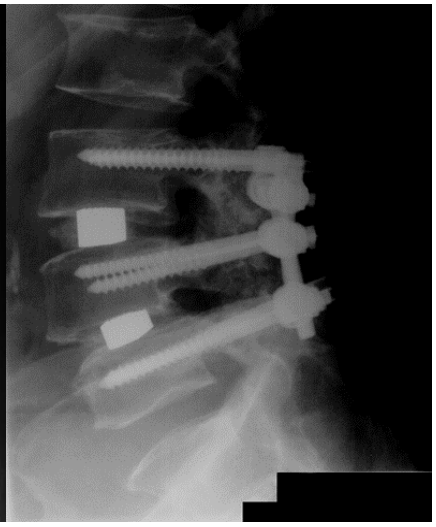
Randomization was done using sealed envelopes with a 20-number-per block randomization. The envelopes were consecutively numbered, thereby assigning a number to each patient in the study. The type of operation remained unknown to both the patient and the surgeon until informed written consent had been obtained from the patient.

Intervention

By the randomization procedure, 49 patients were allocated to posterolateral fusion with titanium TSRH® (Medtronic®) pedicle instrumentation (PLF group) (Picture 2) and 51 patients to TLIF in the form of a tantalum cage (Implex/Zimmer®) placed using an approach lateral to the facet joint (TLIF group) (Picture 3). The anterior interbody fusion device was supported by a posterolateral fusion using pedicle screws (titanium TSRH®, Medtronic®)

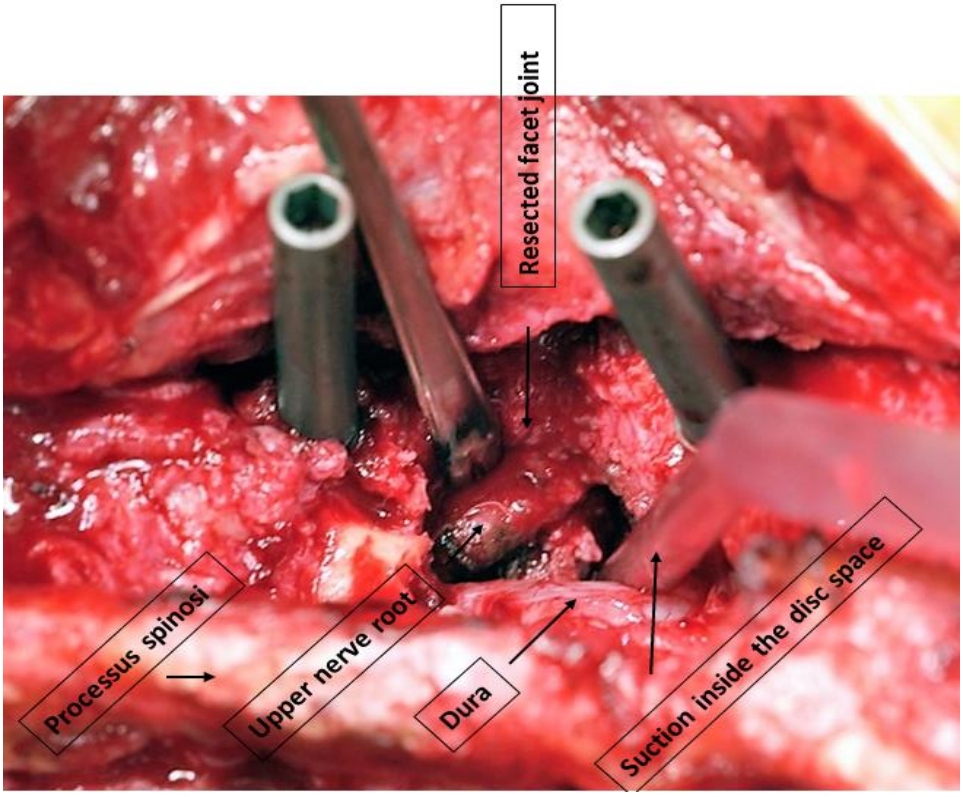


Picture 2. Control PLF



Picture 3. Case TLIF

During surgery, the patients were placed in prone position. Controlled hypotensive anesthesia was used. The patients first underwent insertion of pedicle screws by a midline sub-periosteal approach. When indicated, hemi-laminectomy or laminectomy for neural decompression was performed. In case the patient was randomized to the TLIF procedure, the facet-joint of the



Picture 2. Preparation of operating field during surgery in case TLIF

intended level was identified and the inferior and superior facets were resected in order to gain access to the disc space. The pedicle screws were used to distract the disc space. The upper nerve was identified and protected. The tantalum cage was placed after cleaning the

disc space. Compression over the disc space was done after placement of the cage in order to create lordosis and compression over the implant. Cancellous bone from bone bank femoral heads was used as bone graft and placed laterally on the transverse processes. A careful preparation of the posterolateral region was performed before positioning of the graft. Before that, the decompressed neural structures were covered with a gel-foam (Spongostan®) in order to protect the exposed nerve root. Picture 4 illustrates the preparation and surgical field in a case randomized to TLIF.

Patient-Reported Outcomes Measurement (PROM)

The functional outcome parameters used in this thesis was Owestry Disability Index [49], Dallas Pain Questionnaire [50], SF- 36 [51], and Low Back Pain Rating Scale and Pain Index. [52].

Owestry Disability Index (ODI)

The ODI[49] is a condition specific outcome measure for spinal disorders. It yields an index score which ranges from 0 to 100 with a high percentage reflecting a high degree of disability. It is the most widely used tool and is regarded as "the gold standard" of measuring low back pain. The index includes 10 topics measuring intensity of 1) pain, 2) lifting, 3) ability to self-care, 4) ability to walk, 5) ability to sit, 6) ability to stand, 7) sexual function, 8) social life, 9) sleep quality, and 10) ability to travel. Each topic can be answered with 0 to 5. The scores of all questionnaires are sampled and multiplied by 2 giving the final index. The patients fulfilled the questionnaires independent of the surgeon prior to the operation and after 1, 2, and 5-10 years in the outpatient clinic, still without any influence of the surgeon. The data was collected and calculated by a secretary assistant and pooled into the database. A 10-point difference in ODI is estimated to represent a clinical relevant difference [53]. The ODI version used was 1.0.

Low Back Pain Rating Scale and Pain Index (LBPRS)

Low back pain rating scale[52] was used in both study 1 and 2 in order to assess leg and back pain. The patient grades the pain intensity on a box scale of 0-10, where 10 represents the highest pain and 0 no pain. The scale was developed by Manniche [52] in order to measure the clinical outcome for medical treatment of low back pain. Three categories are used to index the pain experienced in the back and leg using the phrases: Low back pain at the moment, the worst low back pain experienced in the last two weeks, and the average pain level the past weeks. The questions are then repeated regarding the leg pain. Then an 11 point (0-10) numerical rating scale is applied assessing both back and leg pain in three ways: The scores for leg pain and back pain are then summed giving a pain index ranging from 0 to 60. In the papers Study 1, Study 2 and Study 3 for this thesis mostly the pain intensity scale is used, primarily as separate scores for back and leg pain and unless otherwise noted as average pain intensity score. The patients fulfilled the questionnaires independently and prior to surgery, after 1 year, after 2 years (Study 1,3), and between 5-10 years (Study 2). The scoring was done by an independent observer (secretarial assistance) using the above described method.

Dallas Pain Questionnaire (DPQ)

Dallas Pain questionnaire was used as one of the main functional outcome measurements in Study 1 and 2[50]. It has been used extensively in research regarding low back pain in our institution for many years [22, 27, 47, 54-56]. Originally, it was developed and published by Lawlis et al. in 1988 as an instrument to assess the amount of chronic daily spinal pain that affects four aspects of the chronic spine patient: 1) Daily work activities, 2) Leisure activities, 3) Anxiety-depression, and 4) Social interest. The instrument consists of a 16 item visual analog scale (VAS) system: 1) Daily activities including pain and intensity, personal care, lifting, walking sitting, standing, and sleeping; 2) Work and leisure activities including social life, traveling, and vocational; 3) Anxiety-depression including anxiety and mood, emotional control, and depression; 4) Social interest that includes interpersonal relationships, social support, and punishing responses. Each of the 16 items contains its own VAS. The scales are divided into 5-8 small segments in which the patient is asked to place a mark, which indicates the amount of pain between 0% and 100%. For each of these items a score was calculated. The higher score, the greater impact to the patients life. The Danish version of DPQ was used [54]. The patients fulfilled the questionnaires independently and prior to surgery, after 1 year, after 2 years (Study 1 and 2), and between 5-10 years (Study 2). The scoring was done by an independent observer (secretarial assistance) using the primarily described method [50] converted into a Danish version [54]. The instrument has been shown to be internally and

externally valid. The VAS scale provides valid data at different times in percentage along a treatment period [57]. The minimal clinical change that might be significant has been found to be around 20 units [58]. The version used was 1.0, modified to a Danish version [54]

SF-36

The SF-36 is a generic health survey measure [51, 59, 60]. It yields a profile of scores in eight scales covering different physical and mental components of health. The 8 scales are: 1) Physical Functioning (PF), 2) Role-Physical (RP), 3) Bodily Pain (BP), 4) General Health perception (GH), 5) Vitality (VT), 6) Role Emotional (RE), 7) Social Functioning (SF), and 8) Mental Health (MH). The score in each scale ranges from 0 (poorest health) to 100 (best health). Additionally, two summary measures are produced: A Physical Component Summary (PCS) from the first four items and a Mental Component Summary (MCS) of the last four items. The two summary measures are calculated so that the value of 50 is equal to the US population mean. The patients filled out the questionnaires independently and prior to surgery, after 1 year, after 2 years (Study 1 and 4), and between 5-10 years (Study 2). The scoring was done by an independent observer (secretarial assistance) using the primarily described method. The version used was the Danish 1.1.

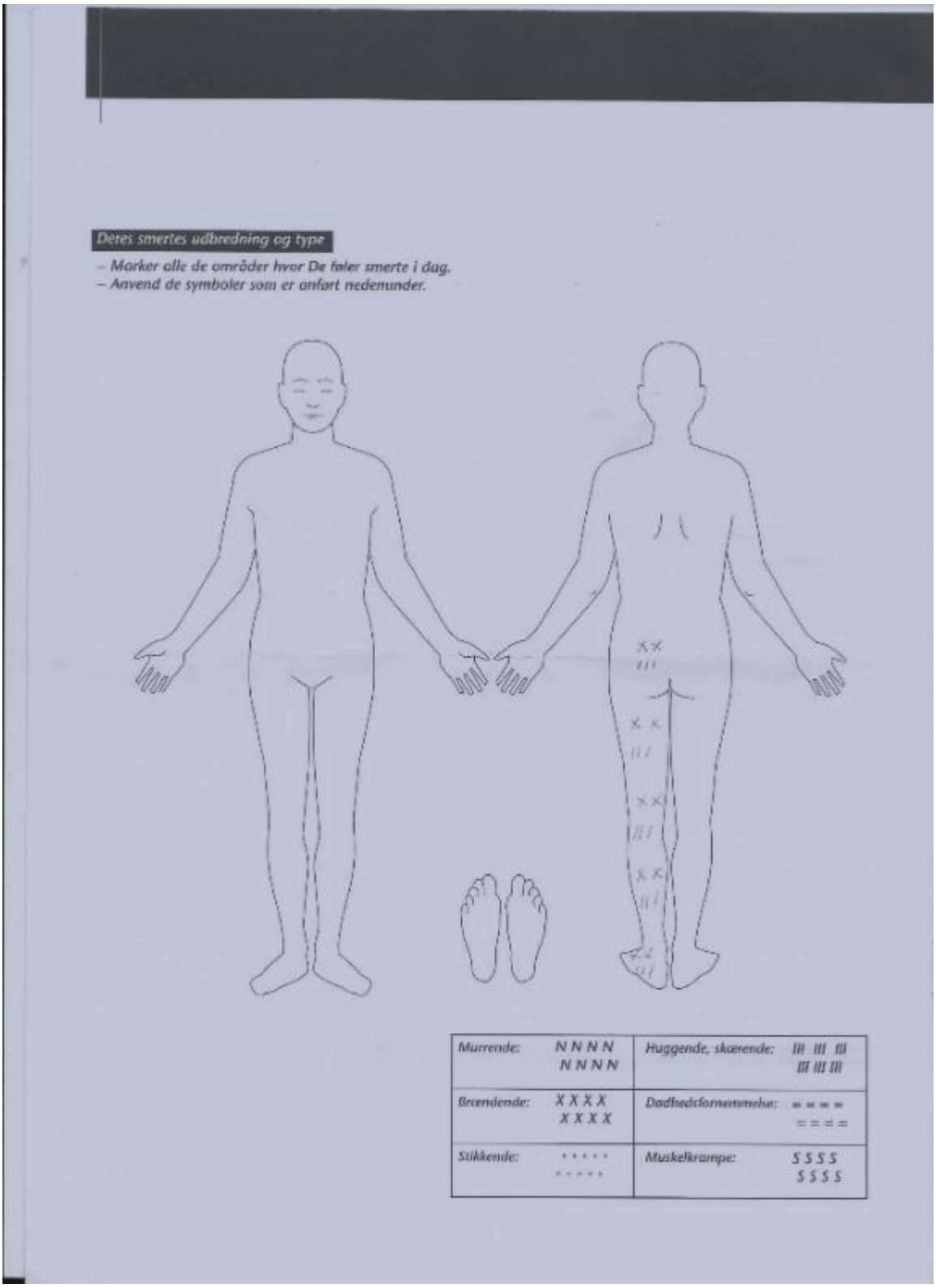
SF-6D

The abbreviated variant SF-6D [61], which can be calculated from SF-36, was used in order to calculate QALY´s for determination of the cost-effectiveness in Study 4.

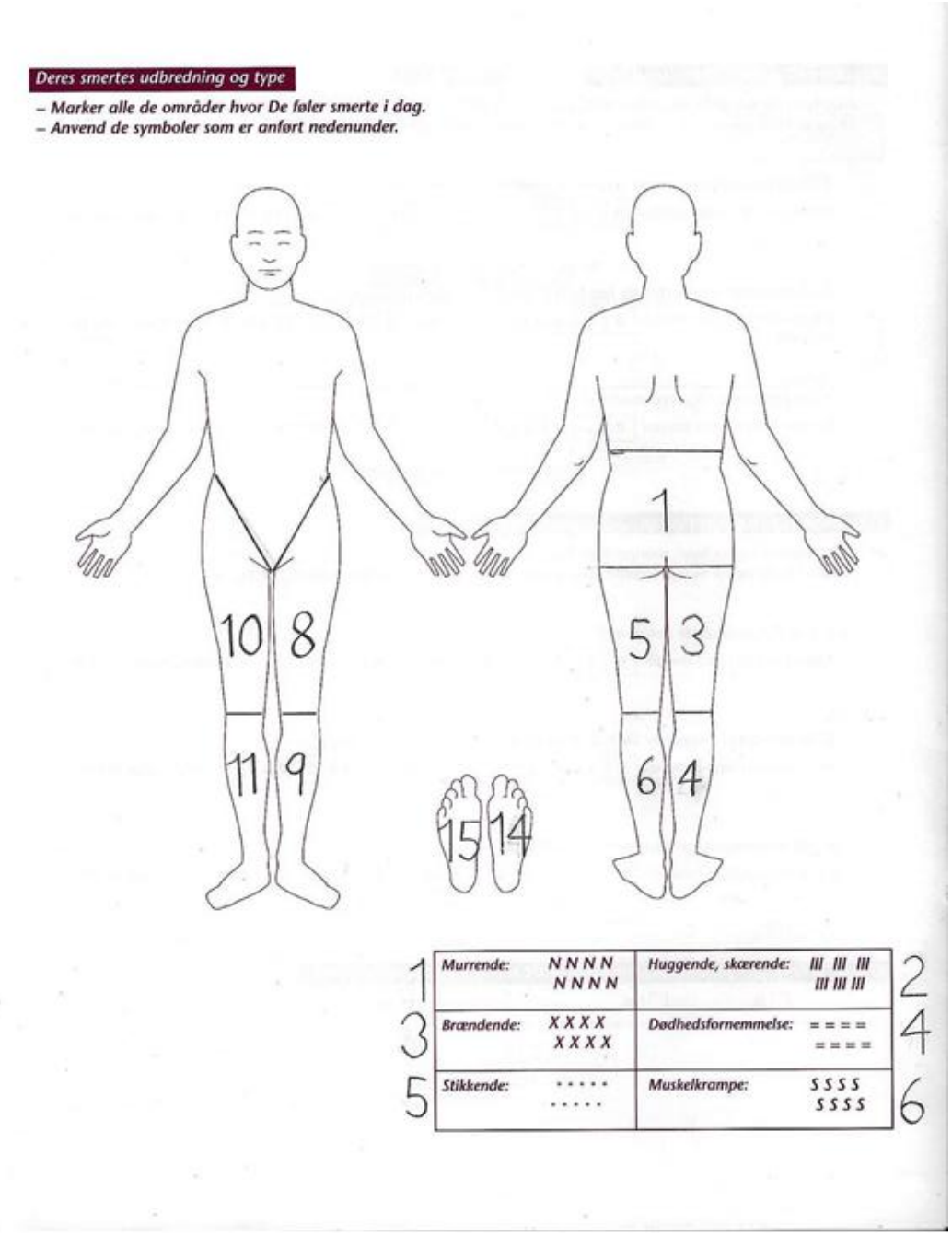
Pain drawings

When filling out the DPQ forms the patients were presented to a front- and-back-outline of a person as well as the area under the feet on the last page of the form (Picture 5). The patients were asked to indicate all their different pains using six different symbols for pain qualities at the present time: 1) dull/aching, 2) burning, 3) numbness, 4) pins and needles, 5) stabbing/cutting, 6) muscular cramps. The pain drawings were divided into 11 areas using transparency overlay and then scored using a combined visual inspection and body region method as first described by Uden and Margolis et al.[62, 63] (Picture 6) with testing for pain presence and pain type in each region. One point was allocated for the presence of one or more pain symbols within each area, resulting in a score range of 0-11. As this scoring method did not include any clinical judgement, the reproducibility was high and the problem of inter- and intra-observer variability very small. Special attention was paid to the side of insertion of the cage in the TLIF group and compared to the presence and type of newly developed pain after surgery. The data was pooled into the 4 areas: posterior thigh, posterior lower leg,

anterior thigh, and anterior lower leg, and the presence of the six above mentioned types of pain in each area was registered. Scoring of the pain drawings in study 3 was done by BG, who was blinded to the two different treatment groups. Data from the pain drawings was pooled and compared between preoperatively and after 1 and 2 years. If present, the leg-pain intensity was registered using a VAS scale 0 (no pain) to 10 (worst) with the leg-pain intensity experienced at the time of examination, with their average perceived leg-pain intensity in the last 14 days and the worst leg-pain experienced in the last two weeks as earlier [52].



Picture 5. Example of pain drawing with S1 affection in left leg



Picture 6. Overlay transparency used for scoring

Other measures

All patients were asked the question “would you undergo the same treatment again, now you know the result?”. The answer could only be yes or no and served as a global satisfaction outcome parameter.

Beside the PROMs, patients were asked about their working status, smoking habits, and the use of painkillers, pain duration, previous spine surgery, and ongoing litigation case.

The indication for surgery, number of levels operated, whether or not additional neural decompression was made, operation time, blood loss, and number of hospital in-days was registered prior to the surgery and immediately after surgery by the single surgeon.

Early and late complications less than 2 years were registered at follow up at 1 and 2 years by the primary surgeon. At 5-10 years, follow-up data concerning complications and reoperations was registered at the visit in the out-patient by the author of this thesis.

Standing X-rays anterior-posterior and lateral views were used to judge the fusion at 1, 2, 5-10 years follow-up by the senior surgeon who conducted the outpatient visit.

Health economic measures

Health economics was calculated by AC under supervision of RS (Study 4) using methods described as evaluation of cost effectiveness alongside a single trail in spine surgery [1] (Figure 2).

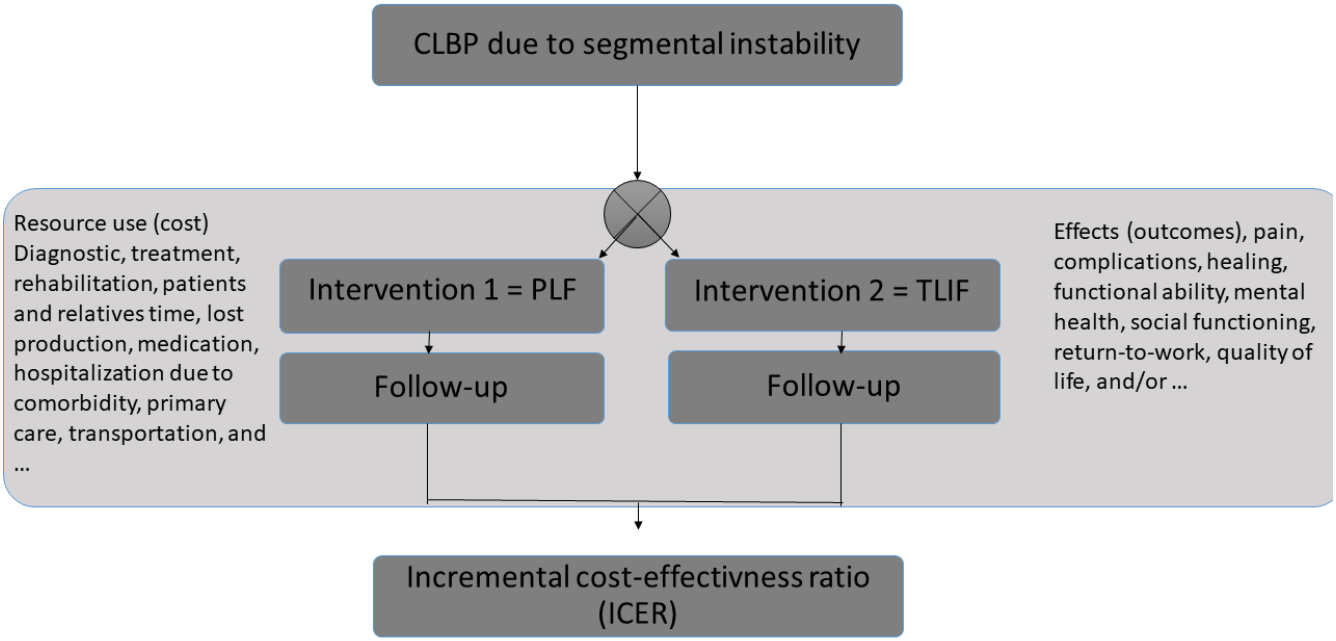


Figure 2. Diagram of resource use and effect measures modified from Søgaard [1]

Cost was divided into cost retrieved from the primary and secondary sector and cost due to sick leave caused by the intervention. The cost in the primary sector included visit to general practitioner, physiotherapy, specialist doctors, as well as cost related to dentistry. Cost retrieved from the secondary health sector was in days spent due to operation, visit to the

outpatients clinic and emergency rooms, including reoperations. The cost from the inability to participate in the normal labor force was also collected.

The cost data related to the surgery was sampled from register data in the National Patient Registry (run by the National Health Service). The cost data for each patient regarding Diagnosis Related Grouping including the reimbursement for the specific treatment was retrieved during the observational period including number of bed days spent inside the hospital [64-66]. The cost spent in the primary sector was collected in the Danish National Health Insurance run by the National Health Service [64, 65, 67] registry for primary care. The tariffs are based on collective agreements between the providers and the National Health Service, which does the reimbursement.

The DREAM data base provided data of persons who were not able to work in the observational period. The production costs were calculated using weeks lost to sick leave among the patients, who were a part of the labor force. Average salaries matched for age and gender were used [65, 68].

Data for effects used in the economical equation ICER (Figure 3) were retrieved using the abbreviated variant SF-6D[61], which was calculated from SF-36[61].

$$\textit{Incremental cost-effectiveness ratio} = \boldsymbol{ICER} = \frac{\boldsymbol{Cost\ TLIF - Cost\ PLF}}{\boldsymbol{Effect\ TLIF - Effect\ PLF}}$$

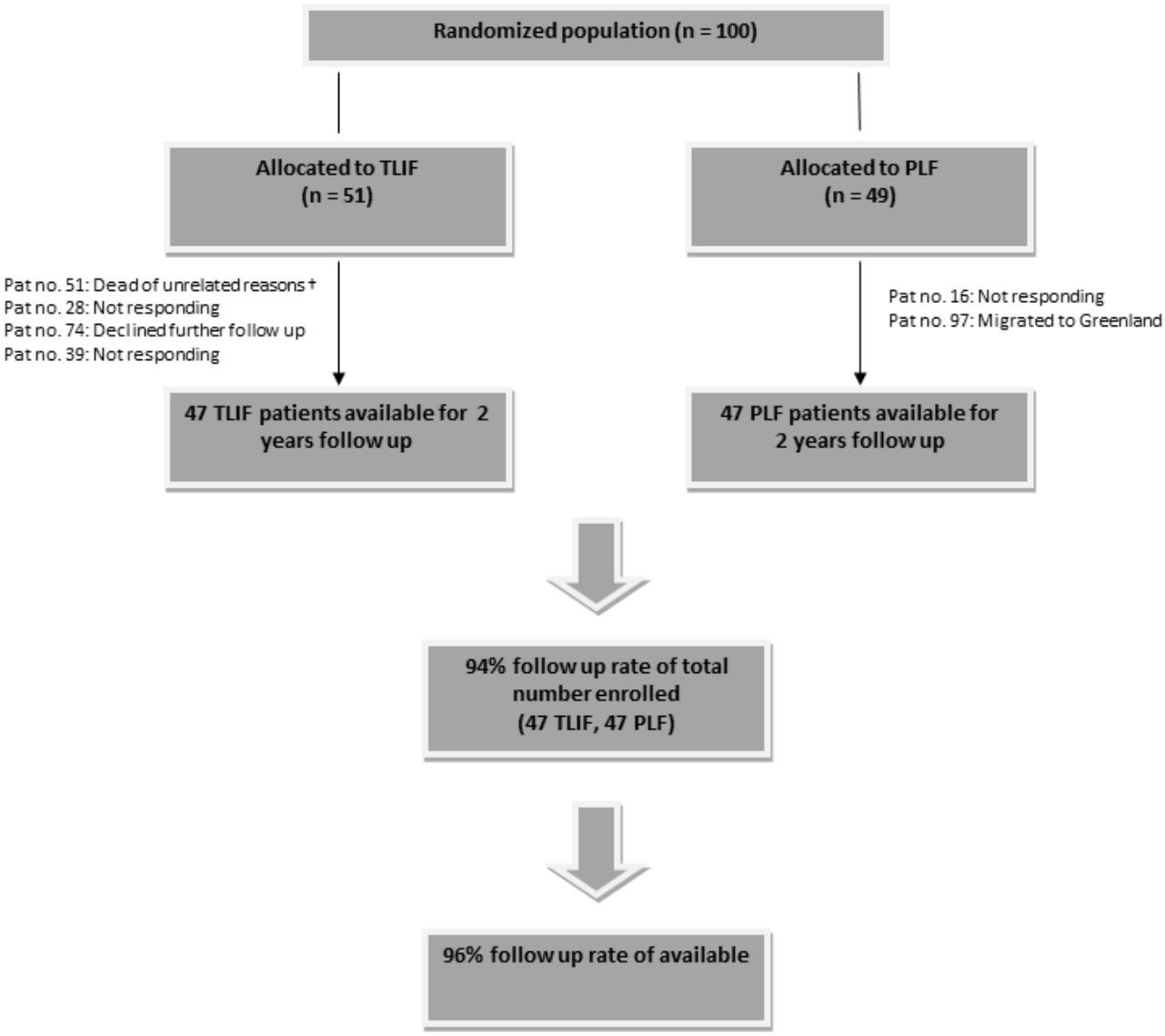
Figure 3. Incremental cost ratio

Baseline Characteristics Study 1, 3, 4

A study population of one hundred patients agreed to participate after having read the written information and signed the written consent. After randomization, 49 were allocated to PLF and 51 to TLIF, see Flowchart 1, Study 1, and Study 3.

The baseline characteristics were equal among the two groups. The groups were identical concerning gender, age, diagnosis, pain duration prior to surgery, number of levels operated, smoking, use of morphine, former surgery, and social status. Exceptions were ongoing compensatory/litigation case and whether laminectomy or laminotomy was done. The latter was due to the different operation techniques, where TLIF always implied partial laminectomy/facetectomy (Table 1).

There were no differences between the two groups concerning DPQ scores, LBRS, ODI, SF-36 PCS, and SF-36 MCS component when entering into the study 1,2,3,4. (Figures 4 and 5).



Flowchart 1. Patients included in study 1, 3 and 4.

	TLIF	PLF	
Sex (female/male)	27/24	32/17	**p=0.209**
Age at surgery (years)	51 (30-63)	49 (25-70)	**p=0.269**
Diagnosis			**p=0.335**
Spondylolisthesis	10	17	
DDD	20	17	
Spinal Stenosis	9	8	
Failed back surgery	12	7	
Pain duration			**p=0.283**
< 1 year	8	3	
1-2 years	14	17	
> 2 years	29	29	
Work status			**p=0.740**
Working	22	16	
Without work	2	2	
Sick-leave	14	15	
Retired	13	16	
Smoking	24 (44%)	21 (46%)	**p=0.904**
Previous spine surgery	14 (27%)	13 (27%)	**p=0.917**
Ongoing litigation case	16 (31%)	7 (14%)	**p=0.042**
Operated level(s)			**p=0.479**
1 level	36	29	
2 levels	14	19	
3 levels	1	1	
Neural decompression			**p<0.001**
None	0	18	
Laminotomy	33	7	
Laminectomy	18	24	

Table 1. Demographics and patient characteristics according to treatment group at baseline. Values are mean (range) or number (%). Litigation case means: insurance, compensatory case etc.

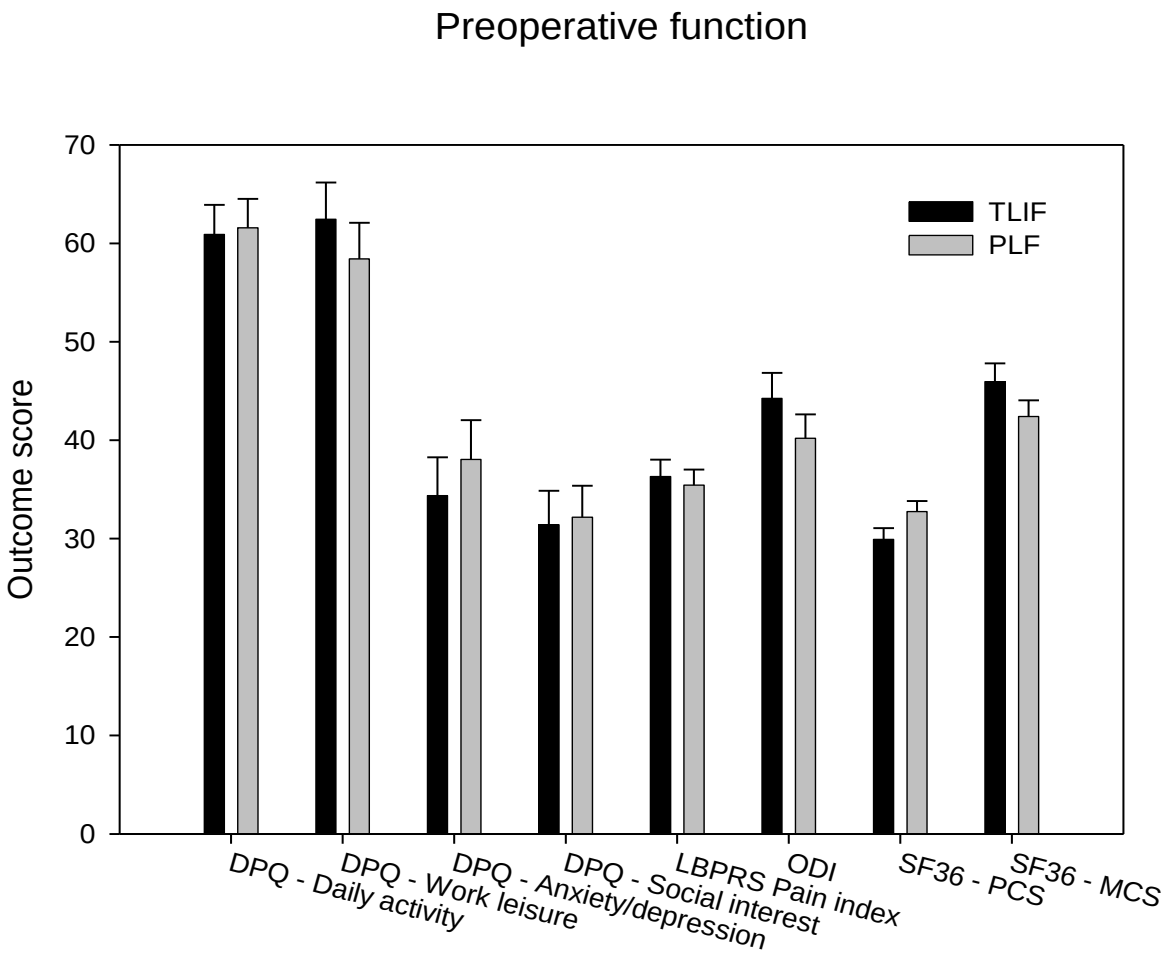


Figure 4. At baseline, preoperatively, there were no difference in PROMs Concerning DPQ, ODI, SF-36, LBPRS pain index between groups. Bars are mean and errors standard error of the mean.

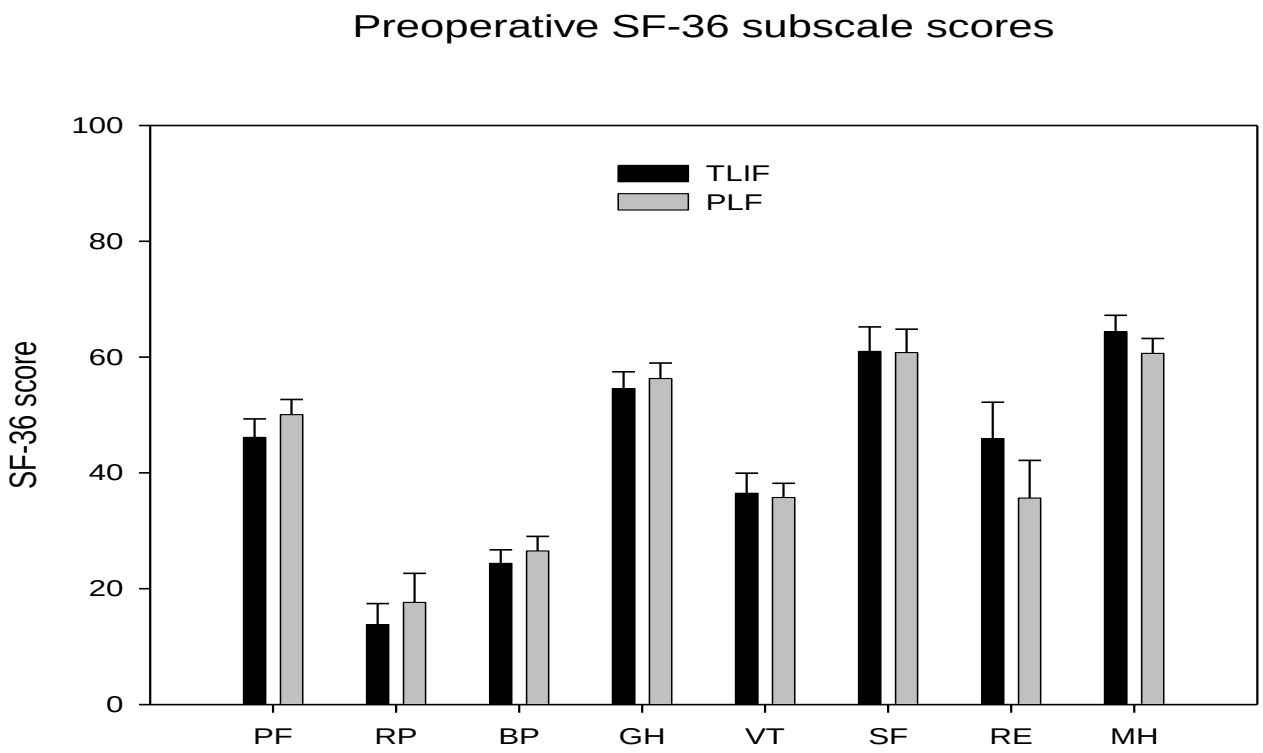


Figure 5. Baseline generic health measurement for all items in SF-36. Physical functioning (PF), Role-Physical (RP), Bodily pain (BP), General health perception (GH), Vitality (VT), Social functioning (SF), Role emotional (RE), and Mental health (MH). With no difference between groups. Bars represent means, error bars standard error of the mean.

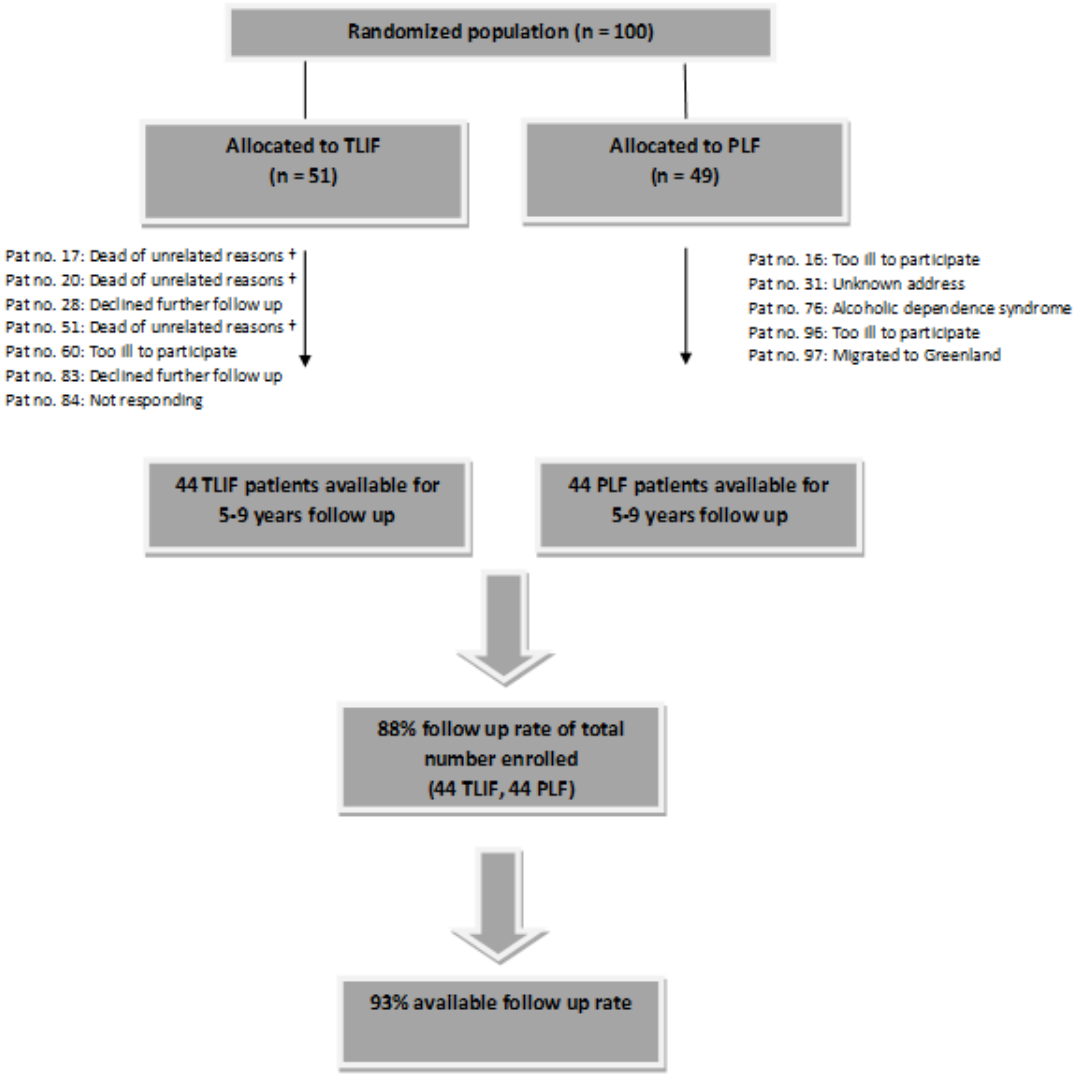
.

Baseline Characteristic Study 2

The material in study 2 consisted of the randomized patients available for follow up 5-10 years after the primary intervention. The patients primarily randomized were asked for a new follow up except for the 3, who had died, one who had migrated to Greenland and one who could not be tracked due to unknown address. The patient flow chart is seen in Flowchart 2. In total 88 patients were eligible for a new follow up. The patients were equally distributed among treatments, and the five blocks used in the randomization period and therefore equally distributed along the entire study period.

At the late follow up all patients had a new X-ray, a clinical examination, and were asked to fill out the patient-related outcome measurements ODI, SF-36, LBPRS, DPQ, and the global question. Pain drawings were not done at this late visit (Table 2).

The patients lost to follow up were equally distributed between the two groups with seven in the TLIF group and five in the PLF group. Their distribution concerning social status, diagnosis, level of surgery, and whether former spine surgery had been done was equal too (Table 3).



Flowchart 2. Patients included and seen in the outpatient clinic at follow up 5-10 years after index surgery (Study 2)

	TLIF	PLF	
Sex (female/male)	22/22	28/16	**p=0.197**
Age at surgery (years)	51 (30-63)	49 (25-70)	**p=0.385**
Age at follow-up (years)	59 (40-73)	57 (34-76)	**p=0.289**
Follow-up length (years)	8.6 (6-10)	8.6 (5-10)	**p=0.796**
Diagnosis			**p=0.563**
Spondylolisthesis	10	16	
DDD	17	15	
Spinal Stenosis	8	6	
Failed back surgery	9	7	
Pain duration			**p=0.451**
< 1 year	6	3	
1-2 years	12	16	
> 2 years	26	25	
Work status			**p=0.349**
Working	21	15	
Without work	2	1	
Sick-leave	12	12	
Retired	9	16	
Smoking	16 (36%)	18 (41%)	**p=0.661**
Previous spine surgery	9 (20%)	13 (30%)	**p=0.325**
Ongoing case	13 (30%)	7 (16%)	**p=0.127**
Operated level(s)			**p=0.386**
1 level	32	26	
2 levels	11	17	
3 levels	1	1	
Values are mean (range) or number (%)			

Table 2. Demographics of patients completing follow-up 5-10 years mean 8.6 years. Values are preoperative numbers unless stated otherwise.

Case #	Group	Sex	Age	Social status	Smoking	Ongoing case	Diagnosis	Previous spine surgery	Level of surgery	Cause of dropout
16	PLF	Female	56	Sick leave	Yes	No	Spinal stenosis	No	L2-L4	Declined participation in follow-up due to poor general condition
17	TLIF	Female	50	Working	Yes	No	DDD	Yes	L4-L5	Died
20	TLIF	Female	62	Retired	No	Yes	DDD	Yes	Other	Died
28	TLIF	Female	61	Retired	Yes	No	DDD	No	L5-S1	Psychiatric disorder. Did not meet for follow-up.
31	PLF	Male	54	Sick leave	Yes	No	Spondylolisthesis	No	L5-S1	Migrated, unknown address.
51	TLIF	Male		Sick leave	Yes	No	Failed back surgery	Yes	L4-S1	Died
60	TLIF	Female	54	Retired	Yes	No	Failed back surgery	Yes	L4-S1	Declined participation in follow-up due to poor general condition
76	PLF	Female	45	Working	No	No	DDD	No	L5-S1	Alcohol abuse, declined participation in follow-up
83	TLIF	Male	46	Sick leave	Yes	Yes	Spinal stenosis	No	L4-S1	Declined participation
84	TLIF	Female	49	Retired	No	Yes	Failed back surgery	Yes	L4-L5	Declined participation
96	PLF	Female	46	Sick leave	No	No	DDD	No	L4-S1	Did not respond
97	PLF	Female	40	Without work	Yes	No	Spinal stenosis	No	L5-S1	Migrated to Greenland.

Table 3. Demographics of patients lost to follow up at 5-10 years. Status at time of surgery.

Statistics

The intention to treat (ITT) principle was used for comparison between the two groups studied. Baseline of the two studied groups was characterized by conventional summary statistics in order to detect any confounding factors relevant to the study population at entry. The confounding factors found relevant for the study population were sex, age, sub-diagnosis, pain duration, work status, former spine surgery, number of operated levels, ongoing compensatory/litigation case. The analysis used for comparison between the two groups were non-parametric tests due to the origin of data used. As seen in the figures of QQ-plots (quantile-quantile plot) of ODI and daily activities (DPQ) the data was not normally distributed, and due to that nonparametric testing was chosen (Figure 6 and 7).

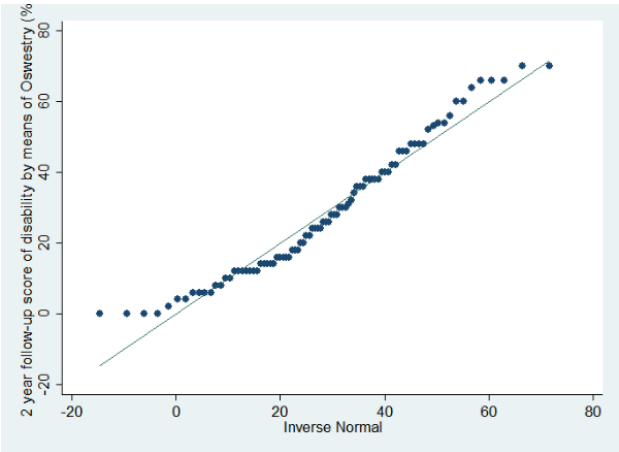


Figure 6. Quantile-quantile plot of ODI values at 2 years follow-up

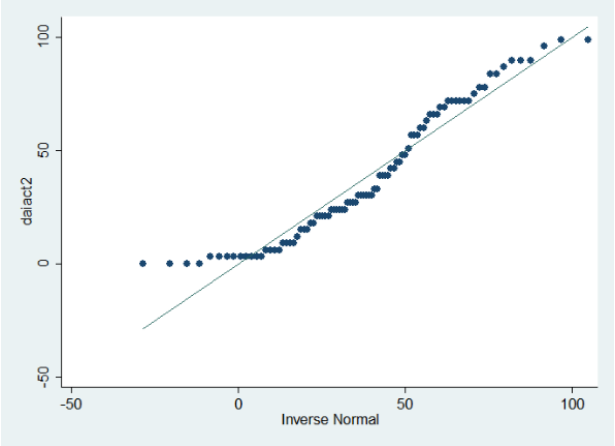


Figure 7. Quantile-quantile plot of daily activities (DPQ) at 2 years follow-up

The points fall along a line in the middle of the graph but curve off at the extremities. Normal Q-Q plots that exhibit this behavior usually mean that data have more extreme values than would be expected if they truly came from a normal distribution.

Level of significance was set to 0.05 (two sided testing). The primary effect measurements were data collected in outcomes measurements, which are ordinal variables, so Wilcoxon/Mann-Whitney rank sum test was used for testing for unpaired data in study 1,2,4. Wilcoxon signed rank sum test was used for paired data in study 1 and 2. Missing data at 5-10 years follow up was equally distributed among the two groups (p=0.5684). Their specific 2 years data regarding ODI, daily activities, work-leisure, anxiety/depression, social interest, and LBPRS was tested against each other. No statistical differences were detected between datasets in outcome scores at that time point, and due to that it was concluded that the problem of missing data did not affect the result, and the analytic strategy was based on cases completing the outcome scores. Chi-square statistics was used for testing the equality between groups. In Study 3, where the data sampled was of both binominal and rank ordinal nature, Chi-square test was used for unpaired data. Correction for ties was done, if necessary, in each data calculation. In Study 4, the cost and outcome measures were skewed, and the calculation was based on arithmetic means with bootstrapped standard errors such that conventional 95% confidence intervals could be estimated. Since data collected in Study 4 came from two different sources, i.e. the RCT and register data, missing data was imputed by

the mean of each randomization group. A non-response analysis was done and no difference between groups were detected. An though-provocative information from the non-response analysis was that no responders of ODI and SF-6D had generated lower production losses. The extent of this was equal between the two groups. A scatter plot of ODI versus operation date was done to test for inferiority caused by performance bias in the study period (Figure 8). All calculations were done using Stata for different Windows versions.

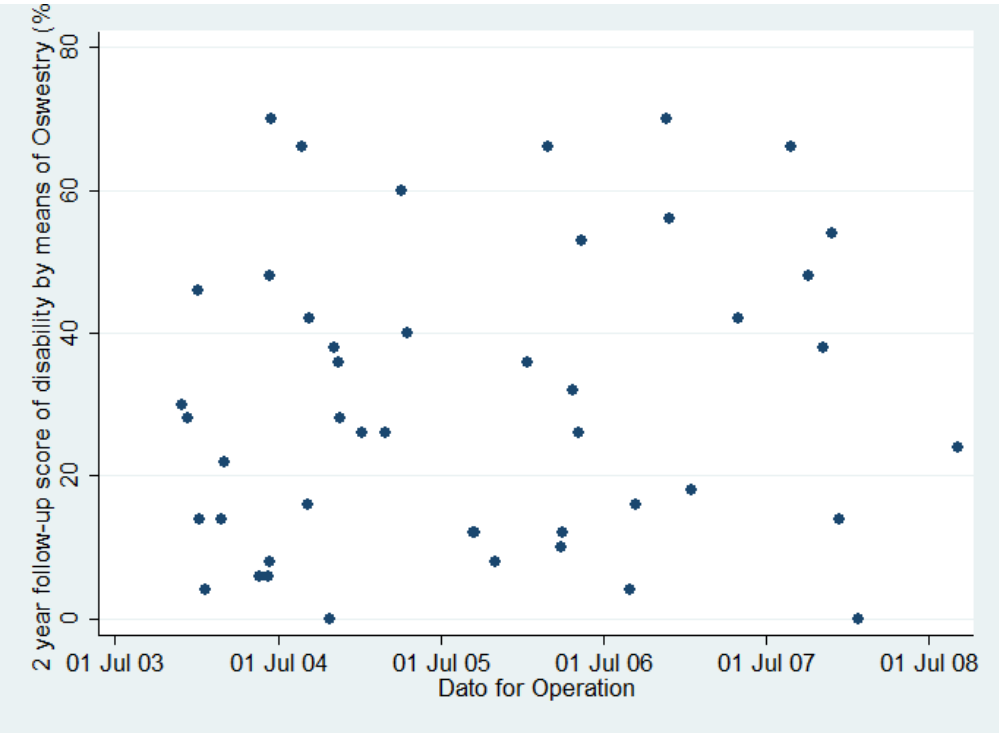


Figure 8. Scatter plot of ODI score at 2 years follow for patients operated on with TLIF. No sign of performance bias due to learning curve during the study period.

Ethical issues

All studies were approved by the Regional Ethics Committee (30030172) and Danish Data Protection Agency (2013-41-1437) and conducted in agreement with the Helsinki Declaration II.

Results

Functional outcome at 1 and 2 years follow up (Study 1)

There were no cross-overs in the study, but two patients in the TLIF group had such a narrow disc space that the cage could not be inserted. Instead, the interbody fusion was made from allograft bone alone, which was impacted between the vertebral endplates through the TLIF approach used for all patients. In one patient from the TLIF group, the operation was changed to an ALIF prior to the operation. This patient were scored after the intention to treat principle ITT.

The results at 2 years follow up did not show any improvement in functional outcome parameters measured at any time point in favor of TLIF compared to PLF. Both groups had significant gains of the fusion procedures compared to the situation prior to surgery (Figure 9 and Table 4).

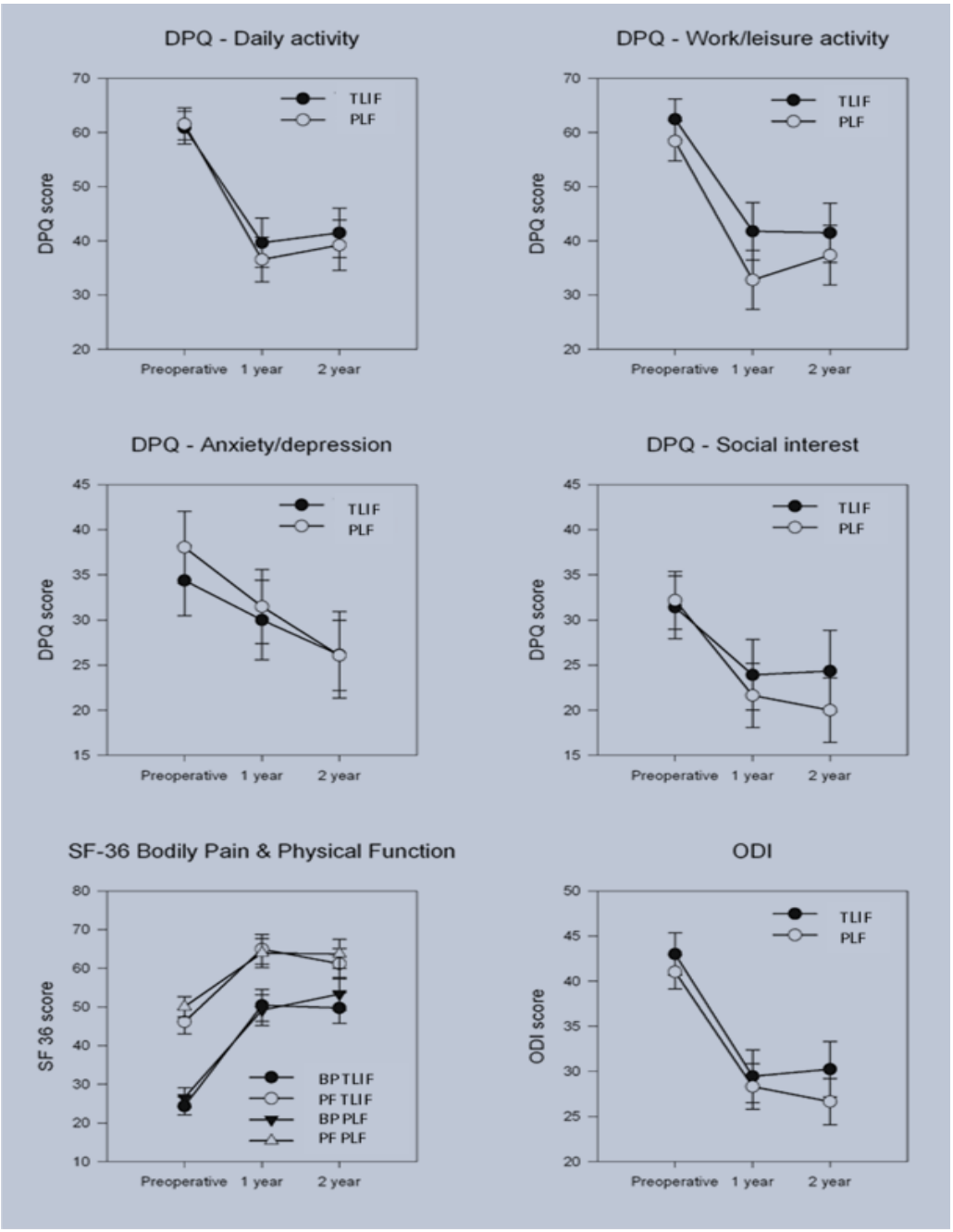


Figure 9. Preoperative, 1 and 2 years follow-up scores of the 4 DPQ aspects, as well as ODI and BP (the bodily pain) and PF (physical function) subscales from the SF-36. There were no significant differences between the two groups at any time points ($p > 0.50$ except DPQ work leisure activity $p > 0.29$). Both groups improved significantly from preoperatively to last follow up ($p < 0.01$, except for DPQ anxiety/depression in the TLIF group $p = 0.03$). Values are mean and errors are standard error of the mean.

A subgroup analysis was performed in order to detect if TLIF had any beneficial effect in some diagnosis subgroup that could support our primary hypothesis in favor of TLIF. No statistical difference was found between treatments in any of the sub-groups included in our study population (Table 4).

The overall satisfaction rate, which served as the global outcome parameter, was answered by 91 patients at the 2 years follow up, and 66 (72%) responded positively, 33/46 in the TLIF group and 33/45 in the PLF group. TLIF was more time consuming than the PLF regarding operation time, which was mean (range) 228 (135-450) min with TLIF versus 171 (100-300) min with PLF ($P > 0.001$), and due to that blood loos was significantly greater in the TLIF group, i.e. mean (range) 775 (150-5000) ml versus 443 (50-1500) ml in the PLF group ($p < 0.001$).

Introducing a new and more complex fusion procedure the complication rate was registered and analyzed in the two groups. A total of 13 % in TLIF group had a complication in comparison to 7 % in the PLF group ($p = 0.205$). The complications in the TLIF group were 1 hematoma, 2 superficial infections, 1 nerve root lesion due to a misplaced cage, 2 dural tears, and one intraoperatively developed pneumothorax. In the PLF group there were 2 hematomas an one dural tear. The fusion rate was equal between the groups, i.e. 94% in the TLIF group and 88 % in the PLF group ($p = 0.31$).

	TLIF	PLF	p-value
Spondylolisthesis			
DPQ Daily activity	40.1 (11.1)	41.8 (9.1)	0.799
DPQ Work/leisure activity	31.1 (10.4)	39.6 (11.5)	0.562
DPQ Anxiety/depression	21.9 (10.7)	18.2 (4.8)	0.808
DPQ Social interest	15.6 (5.8)	21.1 (7.2)	0.974
LBPRS pain index	19.6 (2.7)	17.1 (3.2)	0.503
ODI	25.8 (6.9)	26.7 (4.5)	0.698
SF-36 BP	50.8 (7.0)	57.5 (7.1)	0.509
SF-36 PF	60.8 (8.3)	64.3 (6.6)	0.976
SF-36 PCS	40.0 (3.2)	41.2 (3.5)	0.659
DDD			
DPQ Daily activity	45.5 (7.4)	30.4 (7.3)	0.145
DPQ Work/leisure activity	48.2 (9.2)	32.0 (7.6)	0.155
DPQ Anxiety/depression	29.2 (7.2)	24.3 (6.7)	0.645
DPQ Social interest	25.8 (7.5)	20.0 (5.6)	0.901
LBPRS pain index	26.3 (4.2)	25.0 (5.1)	0.600
ODI	32.4 (4.5)	21.8 (4.0)	0.144
SF-36 BP	48.7 (6.9)	51.9 (8.0)	0.701
SF-36 PF	63.1 (5.6)	70.2 (6.2)	0.431
SF-36 PCS	37.2 (2.4)	42.3 (2.8)	0.184
Spinal stenosis			
DPQ Daily activity	36.4 (9.9)	38.1 (9.8)	0.954
DPQ Work/leisure activity	37.5 (12.5)	35.0 (13.0)	0.942
DPQ Anxiety/depression	15.6 (7.9)	28.6 (9.0)	0.143
DPQ Social interest	18.1 (6.0)	10.7 (7.1)	0.231
LBPRS pain index	23.6 (7.2)	18.1 (4.5)	0.728
ODI	25.0 (6.6)	27.3 (8.5)	1.000
SF-36 BP	52.6 (11.1)	47.4 (11.2)	0.728
SF-36 PF	56.4 (9.7)	62.1 (9.9)	0.560
SF-36 PCS	35.5 (5.4)	38.5 (4.3)	0.729
Failed back surgery			
DPQ Daily activity	39.0 (10.3)	54.4 (12.0)	0.222
DPQ Work/leisure activity	42.5 (12.6)	53.8 (16.6)	0.670
DPQ Anxiety/depression	32.8 (14.1)	45.8 (13.9)	0.551
DPQ Social interest	34.5 (13.5)	28.3 (9.7)	0.912
LBPRS pain index	24.0 (4.7)	25.8 (9.5)	0.957
ODI	34.4 (8.0)	35.7 (5.1)	0.625
SF-36 BP	48.5 (8.0)	53.0 (10.4)	0.404
SF-36 PF	61.6 (11.0)	50.7 (8.9)	0.463
SF-36 PCS	39.5 (4.1)	35.0 (4.4)	0.540

Values are mean (sem)

Table 4. Outcome scores at 2 years follow-up in the two treatment groups stratified according to their surgical diagnosis. Values are mean (standard error of the mean).

Functional outcome at long term follow up (9 years) (Study 2)

The median follow up time was 8.6 years. Since almost ten years had gone by, the mean age had risen to 59 with 60 years in the TLIF group and 58 in the PLF group. Comparing the functional outcome scores between groups no statistical differences were found using the 4 measurements described in the method section (Figure 10).

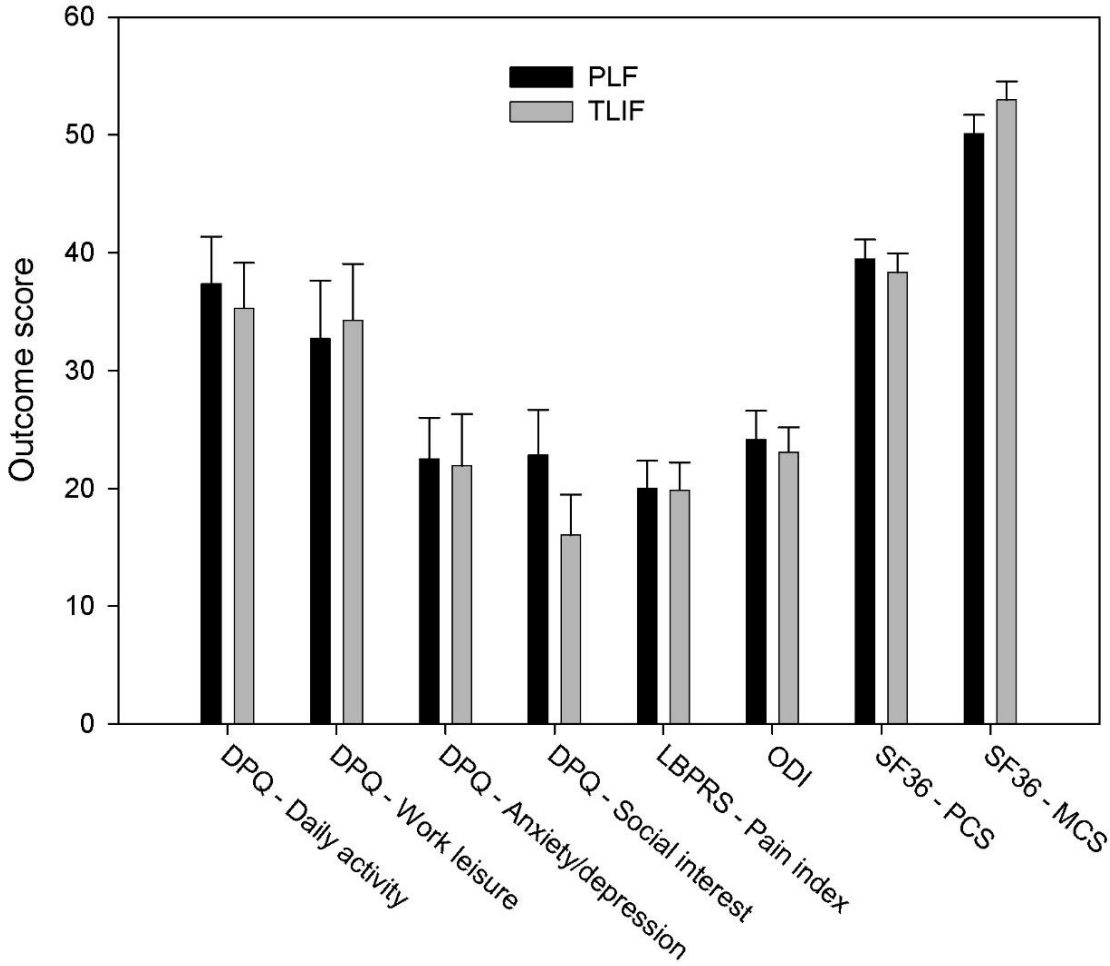


Figure 10. Outcome scores at LTFU in the PLF and TLIF group. None of the differences is significant. Bars represent mean, error bars standard error of the mean.

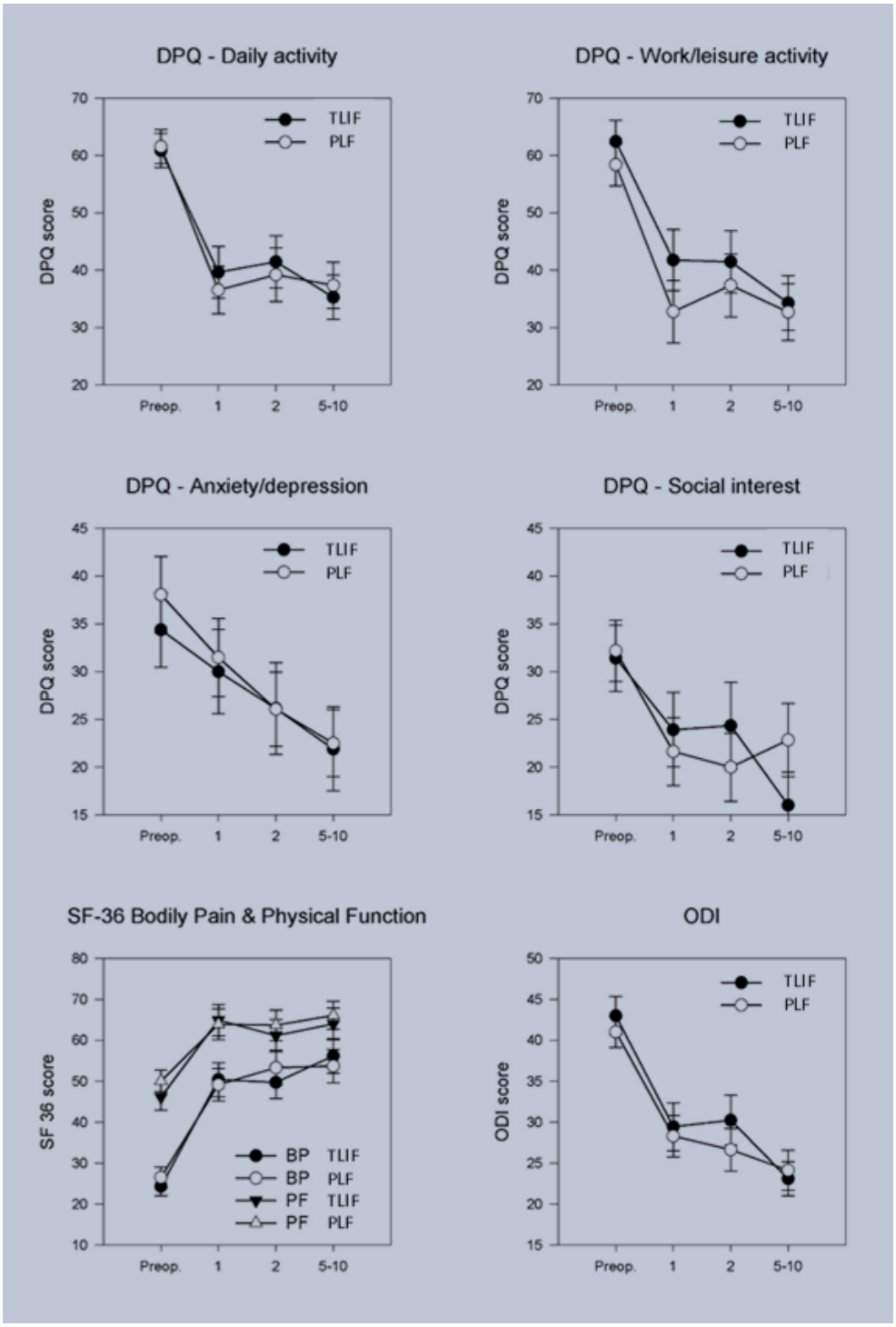


Figure 11 Preoperative, 1, 2, and 5-10 years (mean 8.6) follow-up scores of the 4 DPQ aspects as well as ODI and BP (the bodily pain) and PF (physical function) subscales from the SF-36. There were no significant differences between the two groups at any time points. All long-term scores are significantly better than the preoperative, $p < 0.0305$. Values are mean, and error bars are standard error of the mean.

The primary gain of the surgical intervention stayed significant even after almost 9 years with significant improvements in all used outcome parameters (Figure 11). The overall satisfaction rate, which served as the global outcome parameter, was answered positively by 33 patients in the TLIF group and 34 in the PLF group giving a satisfaction rate of 75 % in the TLIF group and 77% in the PLF group (p=0.803).

At the time point selected some reoperations had occurred in both groups. The early reoperations occurred within the first year after surgery and late reoperation was defined as surgery occurring after 2 years from index surgery. The causes of reoperation are shown in Table 5.

	PLF	TLIF	Total
Early reoperations < 1 year	**4**	**7**	**11**
Hardware loosening	3	3	6
Infection	1	1	2
Early decompression	0	3	3
Disc herniation	0	0	0
Late reoperations > 2 years	**7**	**7**	**14**
ASD	3	4	7
Disc herniation	1	1	2
Hardware removal	3	1	4
Late decompression	0	1	1
I alt	**11**	**14**	**25**

Table 5. Early and late reoperations according to intervention procedure and indication for reoperation. The late decompression and disc herniation was in segments not included in the fusion.

The main reason for late reoperation was adjacent level segment degeneration (ASD) in the disc just above the fusion. There were no differences between the groups. However, patients who underwent reoperation generally experienced poorer functional outcome in all scores at long-term follow-up in comparison to those who only had the index surgery (Figure 12).

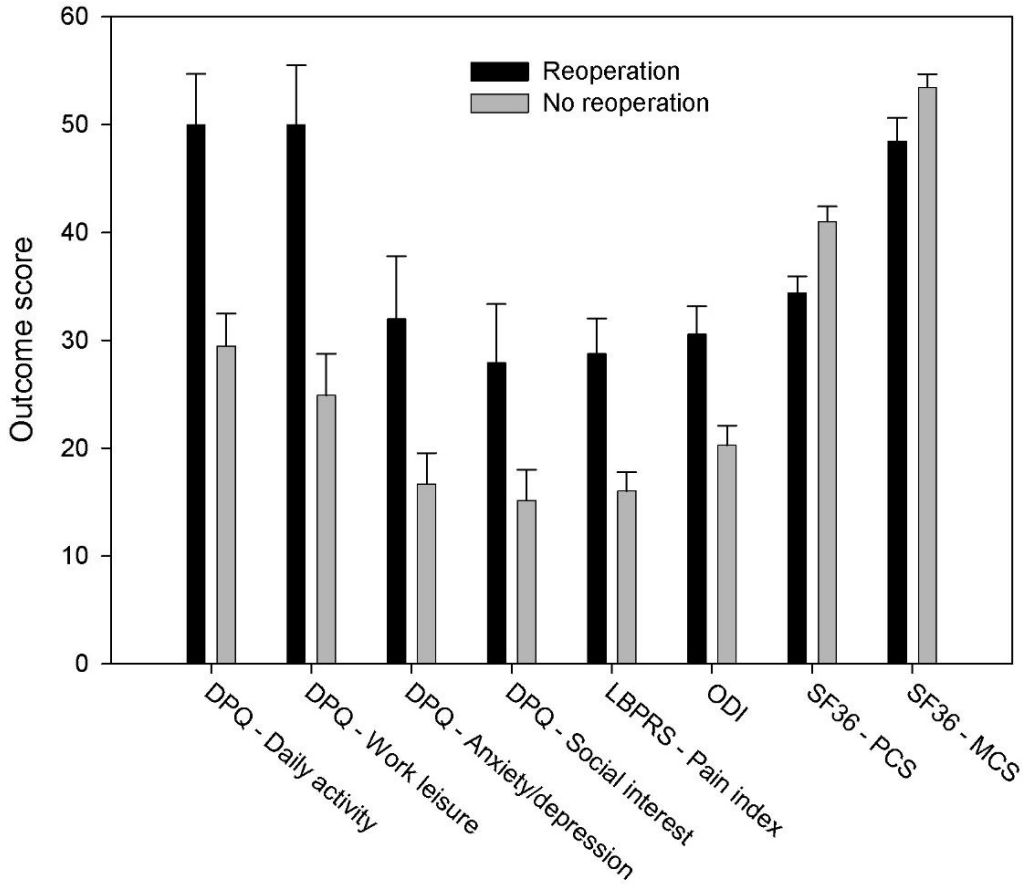


Figure 12. Long-term follow-up outcome scores comparing patients undergoing reoperation during the follow-up period to those only having the index surgery. Bars represent mean, error bars standard error of mean. All differences significant with p<0.03.

Pain drawings and pain intensity measured by LBPRS (Study 3)

Looking upon baseline regarding presence and type of pain in the pooled areas of the pain drawings, there was a skewing in the data material between groups regarding numbness at the posterior thigh. The number of patients who experienced this was higher in the PLF group compared to the TLIF group. In all other types and areas, the baseline was equal between the groups. This skewing was reflected in the data collection at 1 and 2 years follow up (Table 6).

	Preoperatively			1 year follow-up			2 year follow-up		
	TLIF	PLF	p-value	TLIF	PLF	p-value	TLIF	PLF	p-value
Posterior thigh									
Dull/aching	11 (22%)	11 (22%)	0.915	2 (4%)	9 (18%)	0.021	6 (12%)	2 (4%)	0.157
Stabbing/cutting	14 (27%)	11 (22%)	0.564	4 (8%)	4 (8%)	0.953	4 (8%)	4 (8%)	0.953
Burning	10 (20%)	6 (12%)	0.315	1 (2%)	3 (6%)	0.288	1 (2%)	0 (0%)	0.325
Numbness	3 (6%)	9 (18%)	0.055	0 (0%)	7 (14%)	0.005	1 (2%)	5 (10%)	0.083
Pins & needles	4 (8%)	8 (16%)	0.192	2 (4%)	4 (8%)	0.372	6 (12%)	4 (8%)	0.548
Muscular cramps	4 (8%)	5 (10%)	0.680	2 (4%)	2 (4%)	0.967	3 (6%)	3 (6%)	0.960
Posterior lower leg									
Dull/aching	14 (27%)	7 (14%)	0.106	6 (12%)	6 (12%)	0.941	5 (10%)	3 (6%)	0.498
Stabbing/cutting	10 (20%)	7 (14%)	0.479	4 (8%)	5 (10%)	0.680	5 (10%)	5 (10%)	0.947
Burning	9 (18%)	4 (8%)	0.159	3 (6%)	2 (4%)	0.680	2 (4%)	2 (4%)	0.967
Numbness	14 (27%)	12 (24%)	0.736	6 (12%)	10 (20%)	0.239	5 (10%)	8 (16%)	0.332
Pins & needles	13 (25%)	11 (22%)	0.722	5 (10%)	5 (10%)	0.947	5 (10%)	4 (8%)	0.774
Muscular cramps	10 (20%)	10 (20%)	0.920	10 (20%)	4 (8%)	0.099	9 (18%)	5 (10%)	0.284
Anterior thigh									
Dull/aching	11 (22%)	8 (16%)	0.504	5 (10%)	6 (12%)	0.697	10 (20%)	5 (10%)	0.188
Stabbing/cutting	8 (16%)	6 (12%)	0.620	1 (2%)	5 (10%)	0.083	4 (8%)	3 (6%)	0.736
Burning	6 (12%)	5 (10%)	0.803	7 (14%)	3 (6%)	0.205	5 (10%)	1 (2%)	0.102
Numbness	3 (6%)	2 (4%)	0.680	2 (4%)	6 (12%)	0.125	3 (6%)	2 (4%)	0.680
Pins & needles	4 (8%)	6 (12%)	0.463	4 (8%)	2 (4%)	0.428	6 (12%)	1 (2%)	0.057
Muscular cramps	2 (4%)	1 (2%)	0.582	0	0	NA	4 (8%)	1 (2%)	0.183
Anterior lower leg									
Dull/aching	6 (12%)	4 (8%)	0.548	5 (10%)	2 (4%)	0.262	4 (8%)	7 (14%)	0.303
Stabbing/cutting	4 (8%)	6 (12%)	0.463	4 (8%)	6 (12%)	0.463	5 (10%)	1 (2%)	0.102
Burning	5 (10%)	3 (6%)	0.498	4 (8%)	3 (6%)	0.736	2 (4%)	1 (2%)	0.582
Numbness	5 (10%)	7 (14%)	0.491	5 (10%)	7 (14%)	0.491	5 (10%)	3 (6%)	0.498
Pins & needles	4 (8%)	8 (16%)	0.192	7 (14%)	3 (6%)	0.205	6 (12%)	2 (4%)	0.157
Muscular cramps	6 (12%)	4 (8%)	0.548	4 (8%)	0 (0%)	0.045	9 (18%)	4 (8%)	0.159

No patients used more than 2 symbols at one time at any stage

Table 6. Presence and type of pain in the 4 pooled areas. No patient used more than 2 symbols at one time at any stage.

The change in the pain drawings was primarily reflected in the disappearance of pain after surgery. There was a shift from bilateral leg pain to no pain. No difference between groups could be discerned at any time point. Between 1 and 2 years follow up the TLIF group seemed to worsen and the PLF group to improve, still without any significant difference (Figure 13).

The pain intensity was reduced in both groups, but was not different between groups. The reduction in the TLIF group using average leg pain score within the last 2 weeks was 5.1 to 3.5 ($p = <0.01$), and in the PLF group the change went from 5.0 to 2.9 ($p < 0.001$) (Table 7).

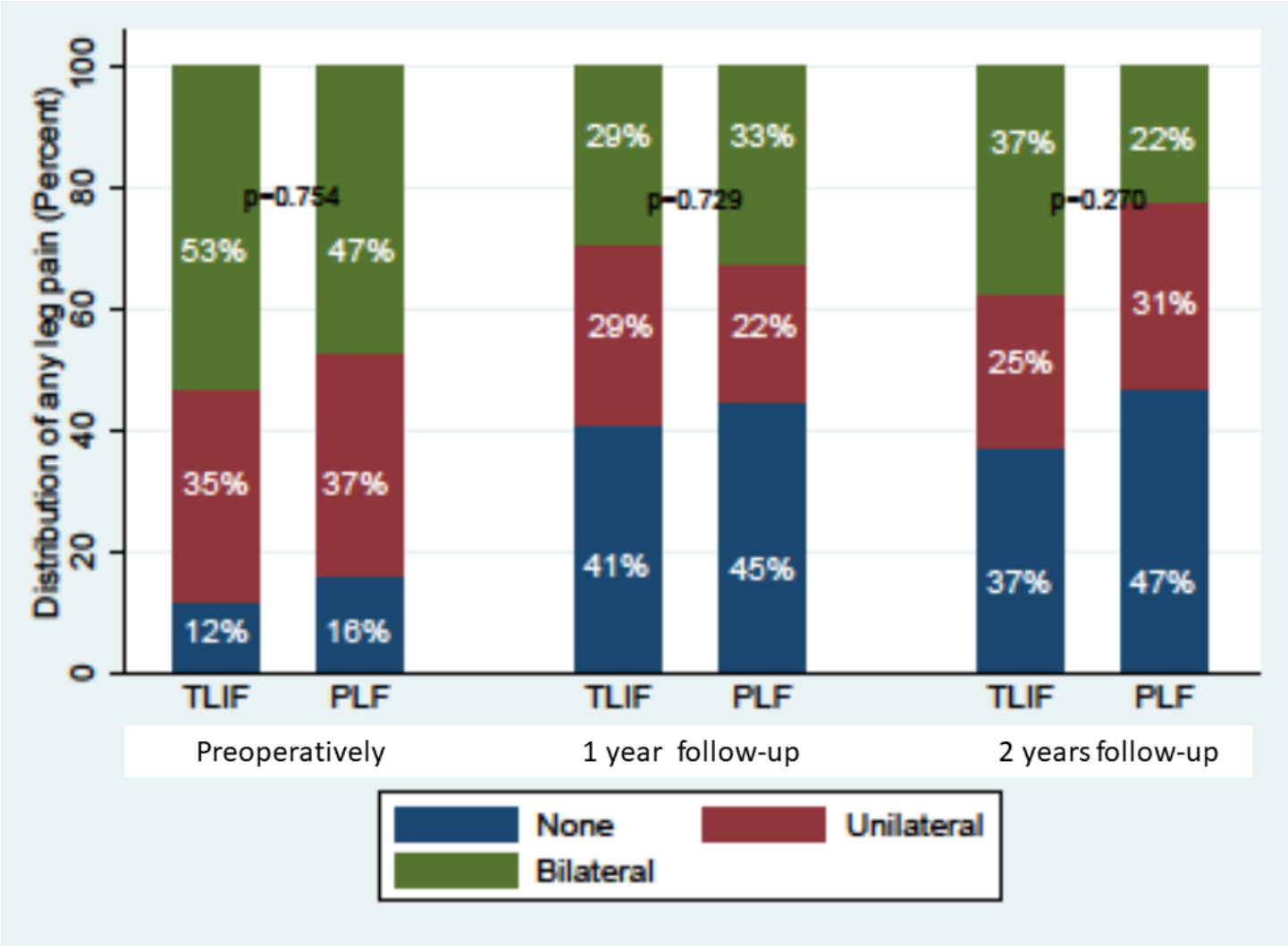


Figure 13. Presence and distribution of leg pain in the whole leg.

Average Leg Pain intensity at 2 years follow up			
	TLIF	**PLF**	
Pre -op	5.1	5.0	p=0.888
Follow -up	3.5	2.9	P=0.214
	P = 0.008	P = 0.0003	
Average Back Pain intensity at 2 years follow up			
	TLIF	**PLF**	
Pre -op	6.1	6.1	p=0.894
Follow -up	4.1	4.0	P=0.774
	P < 0.0003	P < 0.0001	

Table 7. Average leg and back pain intensity measured by VAS scale as average within the last 14 days (LBPRS). Leg pain was significantly reduced in both groups from preoperatively but without any difference between the two intervention groups. Back pain was significantly reduced in both groups from preoperatively but without any difference between the two groups.

	Preoperatively			1 year follow-up			2 year follow-up		
	TLIF	PLF	p-value	TLIF	PLF	p-value	TLIF	PLF	p-value
Back pain right now	6.1 (0.3)	5.8 (0.3)	0.535	3.6 (0.5)	3.4 (0.4)	0.981	3.5 (0.4)	3.6 (0.5)	1.000
Worst back pain within last 14 days	7.8 (0.3)	8.1 (0.3)	0.707	4.9 (0.5)	4.6 (0.4)	0.659	5.3 (0.5)	5.1 (0.5)	0.795
Average back pain within last 14 days	6.1 (0.4)	6.1 (0.3)	0.894	3.8 (0.4)	3.8 (0.4)	0.883	4.1 (0.4)	4.0 (0.5)	0.774
Leg pain right now	5.1 (0.4)	4.3 (0.4)	0.215	3.0 (0.5)	2.4 (0.4)	0.542	3.2 (0.4)	2.5 (0.4)	0.265
Worst leg pain within last 14 days	6.2 (0.4)	6.5 (0.4)	0.657	4.0 (0.6)	3.2 (0.5)	0.470	4.4 (0.5)	3.4 (0.5)	0.137
Average leg pain within last 14 days	5.1 (0.4)	5.0 (0.4)	0.888	3.2 (0.5)	2.7 (0.4)	0.589	3.5 (0.4)	2.9 (0.5)	0.214

Values are mean (standard error of the mean)

Table 8. Average leg pain, worst leg pain, and leg pain right now, at time points preoperatively, 1 year, and 2 years calculated by LBPRS. Values are mean (standard error of mean)

Ipsilateral pain was present in 16% of the patients who had a TLIF cage inserted. Contralateral pain was detected in 14 %. So the side of cage insertion showed no greater amount of pain compared to the contralateral side. The pain totally disappeared in 28 % regarding both ipsilateral pain and contralateral pain. The remainder of the patients continued to be pain free our continued to have pain. At 1 year follow up 5 patients (10%) in the TLIF group had developed new leg pain, 2 patients on the ipsilateral side to cage insertion, 3 patients on the contralateral side. In comparison 7 patients (14%) in the PLF group developed new leg pain ($p=0.491$). TLIF patients developing pain on the side contralateral to cage insertion used the following symbols: Burning (1 patient), numbness (2 patients), pins & needles (1 patient), and cramps (1 patient). The two patients developing ipsilateral pain used: Stabbing, dull, pins & needles, and burning. None of the symbols were used by both patients. The following symbols were used by the 7 patients in the control group: Numbness (5 patients), dull (2 patients), stabbing (2 patients), pins & needles (1 patient).

In the TLIF group, four of the five patients who developed leg pain at 1 year follow-up, did not mark this at their 2 year follow-up. However, another 5 patients marked new presence of leg pain at their 2 year follow-up in a leg marked pain free preoperatively. Two patients marked pain ipsilateral to the side of cage insertion, 2 patients marked pain on the contralateral side, and 1 patient marked the appearance of bilateral leg pain. In the control group only 1 patient

marked pain in a leg marked pain free pre-operatively at 2 year follow-up. (Table 9 and Figure 13).

The pain sensation “pins and needles”, which could be associated with neurogenic pain, was to a slightly higher extent found in the TLIF group almost seen at the anterior part of the leg (Table 8). As seen in Table 8 all different scoring regarding leg pain was in favour of PLF, albeit not significantly. The difference was accentuated between 1 and 2 years.

	TLIF Group		
	Pre-operatively	1 year follow-up	2 year follow-up
No leg pain	6 (12%)	19 (39%)	16 (33%)
Ipsilateral leg pain	13 (25%)	8 (16%)	9 (19%)
Contralateral leg pain	5 (10%)	7 (14%)	4 (8%)
Bilateral leg pain	27 (53%)	15 (33%)	19 (40%)

Table 9. Distribution of leg pain according to side of TLIF cage insertion. Difference in distribution between time point are significant comparing preoperatively with 1 year follow-up (p=0.009) and almost at 2 year follow-up (p=0.082), but with no difference between 1 and 2 year follow-up (p=0.661).

Health economics (Study 4)

The population included in this study was the entire patient cohort included in Study 1, see Flowchart 1. Because data collected in a registry always sums up to 100 %, and because lost values in collected data always is a source of statistical uncertainty, the missing data was imputed by a mean within each randomization group. There was no difference between groups concerning the number of non-responders. In general, the two groups were equal at baseline. The descriptive demographics when entering the study can be seen in Table 10.

	TLIF (n = 51)	PLF (n = 49)	p-value
Age at surgery (mean and range)	51 (30-63)	49 (25-70)	0.27
Male gender	24 (47)	17 (35)	0.21
Work status			0.15
Working	18(35)	11 (23)	
Work shorter hours	4 (8)	4 (8)	
Rehabilitation	0	1 (2)	
Incapacity benefits*	2 (4)	0	
Sick-leave	14 (27)	15 (31)	
Without work	2 (4)	2 (4)	
Early retirement	0	3 (6)	
Retired	11(22)	13 (27)	
ODI (mean and range)	44 (12-86)	40 (18-66)	0.12
SF-6D baseline (mean and range)	0.55 (0.30-0.92)	0.56 (0.40-0.81)	0.63

Table 10. Value are in numbers (percentage) unless other stated. *Incapacity benefits due to back problems.

The PLF group had more visits to the general health practitioner, but not significantly so. In contrast the TLIF group had more visits to the outpatient clinic, but not significantly. The total number of bed days inside the hospital was the same for both groups, but due to the higher

reimbursement of in-days in the DRG grouping system using ICD 10 codes for anterior fusion the cost was 2554 € higher in the TLIF group (not significant). The TLIFs had 4 more days of sick leave than the PLFs resulting in higher societal cost for the TLIF group of 1915 € (Table 11 and 12).

After all data was converted into euros (€) and indexed to price year 2012 a total difference in cost of 4663 € was found between the two intervention groups (Table 12), leaving the treatment TLIF between the first and second quadrant in a cost effective coordinate system, with Δ Effect on the ordinate axis and the Δ Cost on the abscissa. The interpretation of this is, that TLIF should not be introduced as the new standard procedure for surgical treatment of CLBP due to degenerative lumbar disorder, since it cannot be considered cost effective (Figure 14).

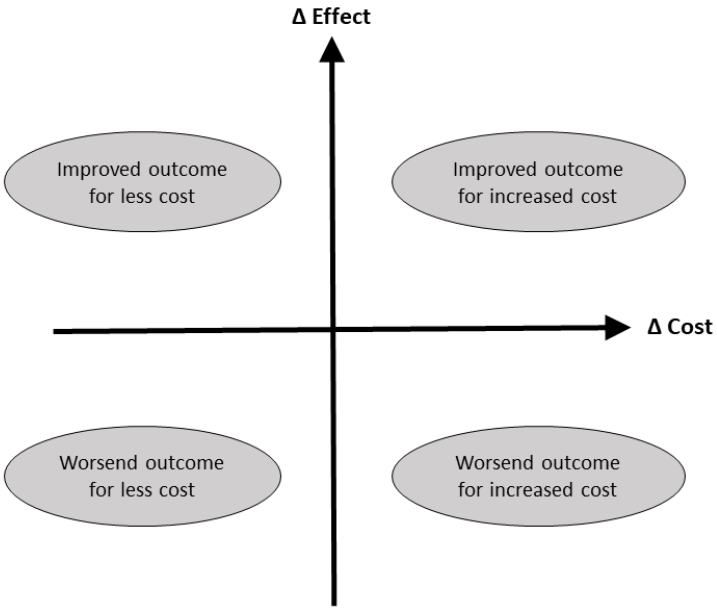


Figure 14. Cost-effectiveness plane

	TLIF (n=51)	PLF (n=49)	Diff.	95% CI for difference (bootstrapped)	p-value (bootstrapped)
Primary health care					
General practitioner	31.96	36.22	-4.26	(-18.54 ; 10.01)	0.55
Medical specialist	1.69	1.61	0.07	(-1.21 ; 1.36)	0.91
Therapist	1.35	2.67	-1.32	(-2.75 ; 0.11)	0.07
Dentistry	2.82	2.98	-0.16	(-0.93 ; 0.62)	0.69
Other	0.06	0.02	0.04	(-0.04 ; 0.11)	0.33
Secondary health care					
Bed days	14.29	14.22	0.07	(-4.45 ; 4.59)	0.91
Out-patient visits	8.73	6.86	1.87	(-1.27 ; 5.00)	0.24
Emergency room	0.10	0.02	0.08	(-0.01 ; 0.17)	0.10
Sick leave					
Weeks	34.20	33.47	0.73	(-13.66 ; 15.11)	0.92

Table 11.Total use of resources during the 2 years follow up. Value are shown as mean units/patient

	TLIF (n=51)	PLF (n=49)	Diff.	95% CI for difference (bootstrapped)	p-value (bootstrapped)
Primary health care					
General practitioner	469	586	-117	(-425.23 ; 192.33)	0.45
Medical specialist	204	146	58	(-112.29 ; 228.22)	0.50
Therapist	126	97	28	(-90.82 ; 146.95)	0.64
Dentistry	116	122	-6	(-42.73 ; 30.46)	0.74
Other	2	1	-1	(-3.30 ; 3.71)	0.91
Primary health care costs	916	953	-37	(-438.51 ; 364.81)	0.86
Secondary health care					
Bed days	22,267	19,713	2,554	(-698.90 ; 5807.18)	0.12
Out-patient visits	1,647	1,428	220	(-426.34 ; 865.36)	0.51
Emergency room visits	13	3	11	(-1.80 ; 22.94)	0.10
Secondary health care costs	23,929	21,144	-2,784	(-771.15 ; 6339.60)	0.13
Sick Leave					
Production loss costs	35,019	33,103	1,915	(-13,253.10 ; 17,085.55)	0.81
Total costs	59,863	55,200	4,663	(-10919.07 ; 20,245.26)	0.56

Table 12. Resources used in Euros. Values are mean cost per patient

Discussion

Discussion Study 1

The testing showed no difference in functional outcome scores at 2 years follow-up, thus the H_0 could not be rejected and H_1 was not accepted. Comparing these results to other similar investigations our findings are in concordance with former and recent publications regarding this topic [22-24, 69], so despite the theoretical advantages of TLIF in comparison with other interbody fusion techniques this was not reflected in the present study population [33]. The use of interbody fusion did not improve outcome after a standard 2 years follow-up. A recent publication, however, has reported significant differences between their two randomization groups in favor of TLIF [28]. Their control group differed from ours, since it was non-instrumented (NPLF). This lead to a higher proportion of crossovers due to frequent pseudo-arthrosis, which was treated with a secondary instrumented fusion. The reoperation rate thus was 18% in their control group and 0 in their TLIF intervention group. As shown in our own study and several other studies, reoperation in spinal fusion leads to less improvement in outcome [70].The authors themselves concluded that the high number of reoperations in their control group might be the reason for superiority in the TLIF group. It is also noteworthy, that their study did not reveal significance between groups in the “gold standard” of low back patient related outcome measurements, i.e. the Oswestry Disability Index (ODI). The level of significance was only reached in pain index and in disability rating index.

The only other RCT study published regarding TLIF in comparison to an instrumented posterolateral fusion was done in a study population of patients with spondylolisthesis grade 1 and 2 [71]. Their results showed significant improvement in the TLIF group, which reached an outcome score as low as 10 in ODI 2 years after surgery versus 68 preoperatively. This result is so different from other published results that doubt must prevail whether it should be believed, primarily due to the tremendous change of 58 in ODI score [71]. In comparison to our subgroup analysis the change in ODI score was only around 15 and in the Swedish spine study the ODI change was between 12 and 14 i.e. from 47-48 to 34-39. In a recently published meta-analysis of TLIF versus PLF with studies from Medline, EMBASE, and Cochrane, only 2 randomized trials were discovered including study 1 from this thesis[72], which was scored to be of high quality. The other RCT included was the one from Jalalpour [28], and that study was graded to be of low quality [72]. The overall conclusion of the review was, that there is no evidence to support a notion that TLIF provides higher fusion rates than PLF and poor evidence to indicate that TLIF might improve clinical outcome. Looking into register data Glassman et al. published data from the Neurosurgical Quality and Outcome Database and found changes around 20 in ODI scores coming from 51.5-53.0 going to 27.2-32.0 in patients who underwent TLIF surgery. The total number of patients included in the cohort studied was 1536. The observation period was only 12 months. Overall, there was a difference between groups in ODI scores between Posterior Spinal Fusion (PSF) and TLIF in favor of TLIF at 12 months, but no significant difference in EuroQOL-5D, and back pain numeric rating scale (NRS)

for leg and back pain. They concluded that TLIF 12 months after surgery showed a greater improvement in ODI scores than PSF, but that the effect was diagnosis specific and only seen in the spondylolisthesis group. The result could not be transferred to other subgroups such as spinal stenosis and adjacent segment degeneration. In contrast to our findings regarding operation time, they found an equal operation time and hence equal blood loss between groups. The explanation for this probably was that the PSF cohort had significantly more levels decompressed than the TLIF group, and no distinction between minimally invasive and open TLIF procedures could be made due to the origin of data (register data). It is well known, that blood loss depends on operation time, and that the risk of complication increases with the degree of technicality [37].

For further comparison to studies with similarities to ours at two years follow up, one must look into the other methods used to achieve interbody fusion in order to widen the discussion and interpretation of our findings. In general, interbody fusion at 2 years follow up had not been able to improve functional outcome [23, 69, 73]. However, trends have pointed in this direction [22]. The results found in our study seem to confirm that interbody fusion does not provide increased outcome scores compared to a standard instrumented lumbar fusion, when tested in a prospective randomized design.

Discussion Study 2

The testing showed no difference in functional outcome scores at 9 years follow-up, thus the H_0 could not be rejected and H_1 was not accepted.

Beside our recently published study [70] no other randomized trials offering long term follow up data (LTFU) of TLIF versus PLF are available, even though it is presently the most used method for lumbar spinal fusion [74]. In this light we are forced to carry the discussion into the field of other interbody spinal fusion procedures such as ALIF and PLIF, were LTFU data are available. The current evidence for interbody fusion is a matter of debate. Recently, Hedlund presented LTFU data from the Swedish lumbar spine study [42]. He was not able to show any difference between conservatively treated low back patients and fused patients in ODI scores, if patients were judged after the ITT principle. Their fusion group consisted of three different arms, one with a non-instrumented PLF group, one with instrumented PLF, and one with instrumented PLF and interbody fusion. The latter was further divided into a PLIF and an ALIF group. In this LTFU no comparison between the patients randomized to the interbody fusion arms versus the patients randomized to non-interbody fusions were done, probably due to insignificant differences in outcome measurements between the different arms, as stated in their primary articles [37, 69]. All functional outcome measurements from all the different fusion arms were pooled, and the overall fusion scores were compared to conservatively treated patients only. Only one other study of high quality, i.e. in the form of

an randomized prospective clinical trial comparing lumbar interbody fusion to instrumented PLF presenting LFTU data, can be tracked and found in Medline[47]. Their results are contradictory to ours, since they found a significant improvement in functional outcome in DPQ, LBRS, SF-36, and ODI with interbody fusion. Looking into the data presented in that article the ODI score was 28 in the interbody fusion group and 40 in the control group. Unfortunately, preoperative ODI scores were not reported in that study. In comparison, our ODI scores were 44 in the PLF´s and 44 in the TLIF´s preoperatively and were reduced to 25 in both groups. Their DPQ scores in daily activities were at entry point 65. At 2 years follow up they had dropped to 40 and 49, respectively. At 7.6 years the ALIF group had dropped further to 31 and the PLF had increased to 55 almost reaching the preoperative value. DPQ scores in our study regarding daily activities went from 61/62 prior to surgery down to 39/41 at 2 years. At the 8.6 years follow up both groups had dropped slightly to 37. Interestingly, both studies were conducted in the same institution, mostly the same surgeons did the surgery, and the indications probably were equal as well. The explanation for the contradictory results may to a certain extent be explained by the severe disability in the control group rather than improvement in the intervention group in the ALIF study [47, 75]. According to the primary article from that study [22] the reoperation rate in the PLF (control) group was significantly higher than that of the interbody fusion group. Since reoperation in spinal fusion surgery is known to lower outcome scores significantly, as seen in our study 2 [70], it can be proposed as the main reason for the difference between groups in the RCT study by Jalalpour [28], as

well as an explanation for the aggravation of disability in the control group rather than improvement in the intervention group of our previous study on ALIF/PLF versus PLF alone [22, 47]. Therefore, from this analysis and due to our own results we conclude that the interbody fusion does not provide better patient related outcomes. However, new RCTs in a multicenter trial might and should provide more information on this topic in the future.

Discussion Study 3

TLIF is presently the most used method to achieve interbody fusion [74, 76]. Since its introduction by Harms [38] nobody has tried to evaluate the risk of the procedure with respect to harm to nerves or creation of permanent radiculopathy, which must be regarded as a possibility due to the exposure of nerves and facetectomy in comparison to the ALIF procedure, which does not include facetectomy. Many TLIF procedures are now preformed with MISS technique, and recent reports claim increased risk of radiculopathy in patients undergoing minimal invasive TLIF [77, 78], which probably is a result of the learning curve in the MISS technique rather than the TLIF procedure per se. [79]. However, the open method needs exposure and handling of the upper nerve, which might be in risk of permanent damage due to coverage of scar tissue or direct trauma during insertion of the cage.

To the best of my knowledge, no studies prior to this have examined this issue in a TLIF cohort in order to detect adverse effects of the method. In order to seek an answer to that question we did a pain drawing study with dichotome variables to test whether the procedure resulted in a higher frequency of nerve pain in the TLIF group in comparison to the PLF group. The method has previously been used in interpretation of spinal pathology [80-82] and has been validated [82, 83]. The study tried to differentiate between presence of pain, type of pain, distribution of pain, appearance of new pain, and the intensity of pain. The study was not able to show a significant difference between groups and these results seem to corroborate other

studies [72, 76, 84]. Therefore, with the aim of achieving interbody fusion TLIF seems to be the method with fewest complications [33, 44, 77, 85-87]. The reason for the slight and insignificant rise in pain in the TLIF group between 1 and 2 years follow up is a matter of speculation. A circumferential stabilization is a more rigid construct, which theoretically might induce increased load to the mobile segment next to the fusion. This might induce an earlier onset of adjacent level degeneration in comparison to that of the normal instrumented fusion (PLF). However, on the other hand the circumferential fusion should theoretically provide better balance of the spine, which certainly should lower the risk of developing ASD. The relation between spinal fusion and ASD is not clarified, and the results are not convincing [21, 88, 89]. The results presented in this thesis [70] did not show a higher amount of reoperations due to ASD in the TLIF group than the PLF group. Therefore, in conclusion, this is just speculative and cannot be regarded as cause of the difference. Overall, the open TLIF method did not reveal un-expected adverse effects in the form of persistent radiculopathy.

Discussion Study 4

The use of health economics is nowadays an important issue and has been proposed to be mandatory before introducing new medical technologies in a patient population in order to gain value for money in a squeezed health care system worldwide. It offers a simple tool for analyzing the incremental cost and incremental effect of a new medical technology in comparison to the gold standard of today [90]. In spinal surgery, it has been used for many years. In 2004 Fritzell did the first study in CLBP patients regarding cost-effectiveness in the Swedish spine study and concluded that surgery might be more cost effective than rehabilitation. The probability of lumbar fusion being cost-effective increased with the value put on extra effect units gained by using surgery. However, they did not compare their different surgery arms, i.e. whether one was more cost-effective than the other [43]. To the best of my knowledge, no one has previously made a prospective randomized trial comparing TLIF to PLF and alongside conducted a matched cost utility evaluation. Our results did not show any significant difference between groups regarding effects or costs. Thus, TLIF, as a new method was not cost-effective over a gold standard PLF. The findings of the study are in accordance with the other studies at two years follow up [91]. However, due to a higher reoperation rate in the PLF group after 3.5 years Bydon et al. concluded that moderate long term cost saving might be the case. Long-term saving was also found in the study from Soegaard [46], who undertook a cost utility evaluation of the study by Videbaek [47] and

found incremental savings of 49306 US dollars pr. gained QALY in favor of circumferential fusion. Therefore, their conclusion was that interbody fusion was significantly cheaper and better in a LTFU. However, they pointed out that the rate of reoperation differed significantly between the groups, which to the point of certainty could explain the difference [46]. Other authors have looked into the cost-effectiveness of TLIF in a subpopulation of CLBPs patients suffering from spondylolisthesis without showing any clear superiority in cost effectiveness of the procedure [45]. Their study consisted of registry data, and it was not possible to distinguish between isthmic and degenerative cases. In addition, the use of either open or MISS technique was not possible to differentiate due to the origin of data. The number of patients in the PSF group was lower than their TLIF group, and they only found improvement in their ODI scores, not in any of their other primary endpoint measurements such as back pain, leg pain, and EQ-5D, which all were prospectively registered. They afterwards calculated the SF-6D from ODI and then had a significance at 12 months. It is also important to recognize, that the PSF group did not reflect weather they were instrumented or not since only fusions greater than 2- level concerning no instrumentation was excluded. Therefore, in general the group of PSFs consisted of both patients non-instrumented and patients instrumented with pedicle screws. The groups were also highly different with only 17.4 % receiving PLF and 82.6 patients receiving TLIF. They were, however, propensity matched in order to deal with the bias such as surgeons skills and preferences, which is a problem in registry studies.

They found equal operation time and blood loss, which probably reflects the use of MISS technique in many of their cases [45]. Overall 12 months of follow up in as population of CLBP patients exposed for surgery is insufficient to uncover any difference in costs and finally in outcomes, which might appear or disappear several years after the index operation [47, 92, 93]. MISS used in combination with the TLIF technique has been found to reduce bleeding, operation time and hospital in-days, and thereby reducing the costs in the MISS TLIF surgery [94, 95]. However, the risk of nerve injury and persistent radiculopathy is claimed and has been shown to be more prominent in patients exposed to MISS surgery [77, 78, 94]. From a societal point of view, one patient with a nerve root injury and persisting radiculopathy will result in a higher use of health care and higher costs over time due to lifelong pain medication and higher use of societal cost due to lost ability for work. A few such cases will with time result in an overall increased use of health related services, which easily overshadows a possible minor gain of one or two days saved in-patient time with the MISS procedure, thus giving less value for money.

General discussion

Since this work was initiated in 2003 new knowledge has come forward. However, to my best knowledge, the studies presented in this thesis are still the first and only to offer an all-around analysis of the TLIF method in a prospective randomized trial in comparison to the gold standard, which is usually thought to be “an instrumented posterolateral fusion” (PLF), with regard to short- and long-term functional outcome. Also, looking into whether the method has negative effects in the form of new radiculopathy due to the cage insertion we find that our studies reveal new knowledge and important information about the method.

The dissertation have tried to measure and cover the most important aspects of a new treatment modality, which ethically should be an obligatory requirement for all new treatments considered for implementation in medical practice before introduced in a larger patient population [96, 97].

This means that the study design chosen must be the most powerful in order to detect any effect of the new method. The trial must try cover the short and long terms effects of the intervention, both positive and negative side/adverse effects. Due to restricted amount of resources in the community, economic considerations must be a part of the entire investigation covering new treatment modalities introduced to a population [1, 10, 97].

The Randomized Clinical Trial (RCT) was chosen in order to provide the strongest statistical evidence free of bias, which was validated in baseline characteristics [97].

Overall, I consider this study valid, since the patients selected were admitted to a university spine unit for judgement of surgical intervention in the form of spinal fusion. The patients selected for fusion were all evaluated and operated on by experienced senior surgeons. Since the study was conducted as a single center study in our spine unit among patients referred to our institution, the surgical indications, the skills of the surgeons, and the homogeneity of the patient population are judged to be uniform, and due to that, the study can be regarded as having a high internal and external validity.

Performance bias was to some extent present, since the patients randomized to the TLIF group always needed an incomplete facetectomy at the side of the cage insertion as opposed to the PLF group. However, it seemed ethically incorrect to do facetectomy in the control group with the risk of permanent neuro-damage, and hence possible adverse side effects were separately investigated in the TLIF group [98].

The outcome measurements chosen were highly validated methods including ODI the gold standard of measurement in judgement of life quality in CLBP patients [49, 99, 100], SF-36, LBRS, and DPQ, and both groups were evaluated equally as expected in RCTs, thus preventing detection bias. The study fulfilled the core criteria of health measurements scales with five domains as proposed by Bombardier with a Health Measurement Scale (HMS) for back specific function ODI, a generic health status SF-36, a Pain scale LBPRS, and finally a global satisfaction question which could be answered yes or no [101].

Validation of LBRS has been done, and actually, the ODI was used for interpreting the validity [49, 52]. DPQ has been practically tested through extensive use in our institution for almost 2 decades [22, 27, 47, 54-56, 102, 103]. However, the latest publication regarding MCID among French CLBP patients has suggested that for a patient to feel well (Patient Acceptable Symptom State, PASS), DPQ scores have to exceed 22, 23, 2 and 10 for daily activities, work and leisure, anxiety/depression, and social interest [104]. None of our fusion groups reached these levels. In our institution the Δ values in daily activities has been determined to be 15 as a criterion to judge success or not in spinal fusion among patients having surgery. Validation of DPQ score has, however, not been done since the introduction by Lawlis in 1989 [50] apart from the French, who also determined the PASS [104, 105]. In our material however, the patients who answered “yes” to the global assessment question had scores in DPQ for daily activities, work and leisure, social interest, and anxiety/depression of 29/26/13/18, respectively. This means that even if the PASS cannot be reached, the intervention still might be regarded successful from a pragmatic point of view. Evidently, it is best to feel good, but it is also good to feel better in case you are dealing with a chronic disease, were the ultimate goal seldom can be reached.

With follow up rates of 96% at 2 years, and 93% at the LTFU, the possibility of bias due to dropouts can be dismissed as an issue, which needs no further discussion, in contrast to that of the few other LTFUs of RCTs published in this field [40, 42, 47, 70].

The length of follow-up was both short 1 and 2 years, which is normal standards for RCT in this field, and long-term LFTU of 9 years complies with the latest scientific trends, which investigate whether the short term effects of the intervention remains or improves, or whether treatments result in undesirable side effects such as adjacent segmental degeneration [40, 42, 47, 88, 89, 106].

The study was conducted without any economical support from the industry and hence free from economical bias, which to a sudden extent have been claimed to drive and fuel the tremendous rise in spinal procedures seen during the last two decades[107, 108].This further strengthens the message of our studies.

BMI and the amount of physical exercise, daily or weekly, could have been confounders at the baseline descriptive statistics. However both DPQ and SF-36 contain questions regarding physical function, and since no difference in those were present, this probably is not the case.

Standing today with the results from this prospective randomized trial, initiated 14 years ago, one might be a little bit disappointed in a hunt for significance! However, after a while, I tend to regard the results as tremendous. We actually succeeded in doing an investigation all around this new method, which we believed could be the solution for fusion procedures in CLBP and degenerative lumbar disorders when the study was started. Today we know that this was not the case, and many new studies point in the same direction [21, 22, 41, 42, 109], although this can be a matter of debate [28, 47]. Therefore, our results seem plausible in

comparison to those of other studies of equal quality. The strength of this thesis is that it covers all aspects of this new method. Short-term follow up, long-term follow up, the adverse side effects, and cost-benefit of the procedure - does it give value for money? And in that context, it gives possibility for an overall evaluation of this new medical technology, which may lead to a change in behavior, due to lack of evidence of its superiority, despite its present status as the most used method worldwide to accomplish interbody fusion [76].

When all this is concluded, one must remember that any conclusions always can and to a certain extent should be challenged. Since the study was conducted from normal standards, with a confidence level of 95% and a power of 80%, a priori statistics leave us with the possibility of 20% having overlooked a real difference between groups in functional outcome in favor of TLIF being superior. By doing a retrospective power calculation from the data collected in our study with a standard deviation of 15 in ODI, (which was the SD of ODI found in our material) and a delta value of 10 points, setting the alpha to 0.05, and doing a two-sided testing, our study sums up to a power of 0.92. This means that the actual risk of doing a type II error, i.e. rejecting the possibility of TLIF being superior to PLF, is only 8.0 % in this material, which further strengthens the conclusions.

The interpretation of this thesis together with previously published literature of high quality should lead to a change in the interpretation of interbody fusions procedures and properly

over time a decrease in the number of these procedures done worldwide. [19, 23, 40-42, 76, 93, 109-115].

When this study was started, the protocol was made in accordance with the standards of that time. However, nowadays and around the start of the study the standards have changed in order to increase the transparency, reliability, and validity of RCTs according to the Consolidated Standards of Reporting Clinical Trials (CONSORT) [116]. These new standards from CONSORT have left this study with a weakness, since the exact number of patients eligible for study participation is not known and cannot be tracked. Individuals excluded because they did not meet the inclusion criteria, declined to participate, or patients not included due to other reasons were not registered separately. This is of course a limitation and in that sense a problem regarding the external validity of the study conducted. However, this study was not, as mentioned earlier, supported by the industry, and in that sense the including surgeons were not biased during enrollment of patients. Besides that, the patients were not able to decide themselves whether they could receive the new TLIF method or not. If they prior to enrollment disagreed to participate in the study, they were offered our standard treatment, a PLF, and due to that, they did not have the opportunity to choose among treatments groups. Certainly, today I would have registered these data during the study period continuously. In hindsight, everything seems clear. However, looking into the checklist from CONSORT 2010, we think that beside the above mentioned limitations, the

results presented in our thesis overall meet the criteria for reporting parallel group randomized trials in accordance with the applicable recommendations from CONSORT as the gold standard of today [116].

In this light I proudly, but humbly, present this contribution to some of the questions, which need an answer in order to improve treatment in the huge patient population suffering from CLBP due to degenerative lumbar disorder.

Conclusions

Conclusion Study 1

In conclusion, the data collected in this study shows that TLIF in a short-term standard follow up length of 2 years does not improve functional outcome or fusion rate as compared to a standard instrumented posterolateral fusion (PLF). Instead, it has longer operation time and higher blood loos. Overall, both TLIF and PLF significantly improves PROMS from preoperative status and, hence, fusion in this sense can be regarded as a clinically relevant treatment modality for patients suffering from CLBP due to degenerative lumbar disorder. Because of the longer operation time and higher blood loss, and since no gain in functional outcome is detected, the recommendation, which can be drawn from this study, in case surgical treatment for CLBP due to degenerative lumbar disorder is indicated, is to choose PLF rather than TLIF.

Conclusion Study 2

Long term follow up data (LTFU) from this study at time point 9 (8.6) years, was not able to prove a significant difference in favor of interbody lumbar fusion with respect to the new TLIF method over a standard instrumented posterolateral fusion PLF regarding PROMS. The ODI scores in both our groups were equal and significantly better than preoperatively. Thus, fusion is a valid treatment modality in CLBP and the effect gained stays for at least 9 years or more. However, reoperations in a population of CLBP individuals lead to significant suffering in a long-term perspective as compared to individuals only exposed to the index surgery. Since ODI scores of 25 in a CLBP patient can be reached by PLF alone with low cost and low risk for the patient, the use of TLIF should not be regarded as a routine method in this patient population.

Conclusion Study 3

With regard to the results collected from our open TLIF surgery, TLIF as a method can be looked upon as a safe method of achieving global fusion. So the conclusions, which can be drawn from our study, are that a slightly higher percentage of patients seems to develop leg pain 2 years after TLIF surgery, but the side of cage insertion does not seem related to a significant increase in the occurrence of radiculopathy or radiculitis, since patients had significant reduction of leg pain at the ipsilateral side of cage insertion and a the slight tendency to new leg pain with time were in both legs.

New literature concerning the TLIF method used as a MISS technique has shown different results with more nerve root injuries and radiculitis at the ipsilateral side of operation, which indicates that if TLIF is to be performed, an open procedure should be preferred judged by our results and the existing literature.

Conclusion Study 4

The TLIF procedure was not able to demonstrate improved outcome, and hence no economic benefit after 2 years was found. The TLIF method in that sense did not give value for money with increased health measured by QALY´s (health –adjusted-life-years). The method demanded more costs during hospital stay and no gain in production loss over PLF. So overall, the competition between interbody fusion and posterolateral fusion did not turn out in favor of TLIF, from a societal point of view.

Suggestions for future research

If we would have wanted to improve our power to 99 % in order to minimize the risk of a type 2 error, the number of patients included should have been 83 in each group. This calls for multicenter studies in order to gain results within a reasonable time. However, the internal validity in multicenter trials might be influenced by different referral patterns in the patient populations, differences in culture, and different interpretation of diagnosis among the medical staffs.

Another possibility could be register research, which could deliver enough patients over time. The Danish Spine Registry has recently become nationwide, and spine registries of Norway and Sweden have been running for several years and supply the spine community with new and important knowledge. One must, however, remember that register research always suffers from poor validation of data quality and selection of patients.

The future will call for clinical trials in which inclusion of controls without surgical intervention is mandatory in order to test the effect of new spine technologies, since new controversies have arisen after presentation of LFTU data from the three randomized clinical trials comparing rehabilitation to surgery in a population of CLBP [17, 18, 40-42, 69, 109].

One must remember that much of the research done is driven by economic incitements from the industry, and future studies should be done without the influence of economic interests.

Therefore, governments or other independent sources such as the EU committee or independent spine organizations should support studies in order to direct the research into a non-commercial direction.

Recently, minimally invasive surgery-especially in the form of TLIF has come along, and its place is still debatable since there have been a rise in nerve root injuries concomitant with the introduction of these different MISS techniques. RCTs comparing standard procedures are urgently needed. [77, 78, 94].

If a method is put into a broad patient population, the positive effect might be hidden, and the method might be unjustified, even if it had a positive outcome in a smaller population at the right indication. Looking at the front page of this dissertation, and knowing the classification of lateral stenosis [2, 117], it seems at least theoretically sound to assume that interbody fusion releases the impinged nerve structures in zone 2 of lateral recess stenosis, and a study put up in this little subgroup of patients should be tried in a multicenter set up.

However, it is always important to seek knowledge in research done in the strongest set up possible [114]. Thus, experiments based on logic theory might not fall out as excepted.

References

1. Søgaard, R., M.E. Kløjgaard and J. Olsen (2010), *Methods for cost-effectiveness evaluation alongside trials in spine surgery*, Centre for Health Services Research and Technology Assesment: University of Southern Denmark.
2. Kirkaldy-Willis, W.H., J.H. Wedge, K. Yong-Hing, et al. (1978) *Pathology and pathogenesis of lumbar spondylosis and stenosis.* Spine (Phila Pa 1976), **3**:319-28.
3. Koes, B.W., M.W. van Tulder and S. Thomas (2006) *Diagnosis and treatment of low back pain.* BMJ, **332**:1430-4.
4. retningslinje, N.k. (2016) *https://www.sst.dk/da/udgivelser/2016/nkr-laenderygsmerter*.
5. Andersson, G.B. (1999) *Epidemiological features of chronic low-back pain.* Lancet, **354**:581-5.
6. *Low-back pain : Frequency, management and prevention from an HTA perspective*. Danish Health Technology Assessment ; 1(1). 1999, Cph.: Danish Institute for Health Technology Assessment, National Board of Health. 106.
7. Koch, M.B., M. Davidsen and K. Juel, *De samfundsmæssige omkostninger ved rygsygdomme og rygsmerter i Danmark*. 2011, Kbh.: Statens Institut for Folkesundhed. 43.
8. Damsgaard, J.B., *A struggle for recognition and meaning : a phenomenological-hermeneutic study of the illness trajectory of patients undergoing spinal fusion surgery : PhD dissertation*. [Publikationer fra Institut for Sygeplejevidenskab, Aarhus Universitet. 2016, Aarhus: Aarhus Universitet, Section for Nursing Science. 132.
9. Deyo, R.A., A. Nachemson and S.K. Mirza (2004) *Spinal-fusion surgery - the case for restraint.* N Engl J Med, **350**:722-6.
10. Søgaard, R., *Lumbar spinal fusion and postoperative rehabilitation : a health economic evaluation : PhD thesis*. 2007, Århus: Faculty of Health Sciences, University of Aarhus. 53.
11. Deyo, R.A. and J.N. Weinstein (2001) *Low back pain.* N Engl J Med, **344**:363-70.
12. Frymoyer, J.W. and D.K. Selby (1985) *Segmental instability. Rationale for treatment.* Spine (Phila Pa 1976), **10**:280-6.
13. Adams, M.A. and P.J. Roughley (2006) *What is intervertebral disc degeneration, and what causes it?* Spine (Phila Pa 1976), **31**:2151-61.
14. Edgar, M.A. (2007) *The nerve supply of the lumbar intervertebral disc.* J Bone Joint Surg Br, **89**:1135-9.
15. Barrick, W.T., J.A. Schofferman, J.B. Reynolds, et al. (2000) *Anterior lumbar fusion improves discogenic pain at levels of prior posterolateral fusion.* Spine (Phila Pa 1976), **25**:853-7.
16. Lee, S.Y., N.H. Cho, Y.O. Jung, et al. (2017) *Prevalence and Risk Factors for Lumbar Spondylosis and Its Association with Low Back Pain among Rural Korean Residents.* J Korean Neurosurg Soc, **60**:67-74.
17. Brox, J.I., O. Reikeras, O. Nygaard, et al. (2006) *Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: a prospective randomized controlled study.* Pain, **122**:145-55.
18. Brox, J.I., R. Sorensen, A. Friis, et al. (2003) *Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration.* Spine (Phila Pa 1976), **28**:1913-21.
19. Fritzell, P., O. Hagg, P. Wessberg, et al. (2002) *Chronic low back pain and fusion: a comparison of three surgical techniques: a prospective multicenter randomized study from the Swedish lumbar spine study group.* Spine (Phila Pa 1976), **27**:1131-41.

20. Moller, H. and R. Hedlund (2000) *Instrumented and noninstrumented posterolateral fusion in adult spondylolisthesis--a prospective randomized study: part 2.* Spine (Phila Pa 1976), **25**:1716-21.
21. Ekman, P., H. Moller and R. Hedlund (2005) *The long-term effect of posterolateral fusion in adult isthmic spondylolisthesis: a randomized controlled study.* Spine J, **5**:36-44.
22. Christensen, F.B., E.S. Hansen, S.P. Eiskjaer, et al. (2002) *Circumferential lumbar spinal fusion with Brantigan cage versus posterolateral fusion with titanium Cotrel-Dubousset instrumentation: a prospective, randomized clinical study of 146 patients.* Spine (Phila Pa 1976), **27**:2674-83.
23. Ekman, P., H. Moller, T. Tullberg, et al. (2007) *Posterior lumbar interbody fusion versus posterolateral fusion in adult isthmic spondylolisthesis.* Spine (Phila Pa 1976), **32**:2178-83.
24. Høy, K., C. Bunger, B. Niederman, et al. (2013) *Transforaminal lumbar interbody fusion (TLIF) versus posterolateral instrumented fusion (PLF) in degenerative lumbar disorders: a randomized clinical trial with 2-year follow-up.* Eur Spine J, **22**:2022-9.
25. Maetzel, A. and L. Li (2002) *The economic burden of low back pain: a review of studies published between 1996 and 2001.* Best Pract Res Clin Rheumatol, **16**:23-30.
26. Gibson, J.N. and G. Waddell (2005) *Surgery for degenerative lumbar spondylosis.* Cochrane Database Syst Rev, CD001352.
27. Thomsen, K., F.B. Christensen, S.P. Eiskjaer, et al. (1997) *1997 Volvo Award winner in clinical studies. The effect of pedicle screw instrumentation on functional outcome and fusion rates in posterolateral lumbar spinal fusion: a prospective, randomized clinical study.* Spine (Phila Pa 1976), **22**:2813-22.
28. Jalalpour, K., P. Neumann, C. Johansson, et al. (2015) *A Randomized Controlled Trial Comparing Transforaminal Lumbar Interbody Fusion and Uninstrumented Posterolateral Fusion in the Degenerative Lumbar Spine.* Global Spine J, **5**:322-8.
29. Mummaneni, P.V., S.S. Dhall, J.C. Eck, et al. (2014) *Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 11: interbody techniques for lumbar fusion.* J Neurosurg Spine, **21**:67-74.
30. Briggs, H. and P.R. Milligan (1944) *Chip fusion of the low back following exploration of the spinal canal.* Journal of Bone and Joint Surgery, **26**:125-130.
31. Cloward, R.B. (1953) *The treatment of ruptured lumbar intervertebral discs; criteria for spinal fusion.* Am J Surg, **86**:145-51.
32. Brantigan, J.W., A.D. Steffee, M.L. Lewis, et al. (2000) *Lumbar interbody fusion using the Brantigan I/F Cage for posterior lumbar interbody fusion and the variable pedicle screw placement system - Two-year results from a Food and Drug Administration Investigational Device Exemption Clinical Trial.* Spine, **25**:1437-1446.
33. Humphreys, S.C., S.D. Hodges, A.G. Patwardhan, et al. (2001) *Comparison of posterior and transforaminal approaches to lumbar interbody fusion.* Spine (Phila Pa 1976), **26**:567-71.
34. Brantigan, J.W., A. Neidre and J.S. Toohey (2004) *The Lumbar I/F Cage for posterior lumbar interbody fusion with the variable screw placement system: 10-year results of a Food and Drug Administration clinical trial.* Spine J, **4**:681-8.
35. Greenough, C.G., L.J. Taylor and R.D. Fraser (1994) *Anterior lumbar fusion: results, assessment techniques and prognostic factors.* Eur Spine J, **3**:225-30.
36. Christensen, F.B. and C.E. Bunger (1997) *Retrograde ejaculation after retroperitoneal lower lumbar interbody fusion.* Int Orthop, **21**:176-80.
37. Fritzell, P., O. Hagg, A. Nordwall, et al. (2003) *Complications in lumbar fusion surgery for chronic low back pain: comparison of three surgical techniques used in a prospective randomized study. A report from the Swedish Lumbar Spine Study Group.* Eur Spine J, **12**:178-89.

38. Harms, J.G. and D. Jeszenszky (1998) *Die posteriore, lumbale, interkorporelle Fusion in unilateraler transforaminaler Technik.* Oper Orthop Traumatol, **10**:90-102.
39. Babu, M.A., J.V. Coumans, B.S. Carter, et al. (2011) *A review of lumbar spinal instrumentation: evidence and controversy.* J Neurol Neurosurg Psychiatry, **82**:948-51.
40. Mannion, A.F., J.I. Brox and J.C. Fairbank (2013) *Comparison of spinal fusion and nonoperative treatment in patients with chronic low back pain: long-term follow-up of three randomized controlled trials.* Spine J, **13**:1438-48.
41. Fairbank, J., H. Frost, J. Wilson-MacDonald, et al. (2005) *Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial.* BMJ, **330**:1233.
42. Hedlund, R., C. Johansson, O. Hagg, et al. (2016) *The long-term outcome of lumbar fusion in the Swedish lumbar spine study.* Spine J, **16**:579-87.
43. Fritzell, P., O. Hagg, D. Jonsson, et al. (2004) *Cost-effectiveness of lumbar fusion and nonsurgical treatment for chronic low back pain in the Swedish Lumbar Spine Study: a multicenter, randomized, controlled trial from the Swedish Lumbar Spine Study Group.* Spine (Phila Pa 1976), **29**:421-34; discussion Z3.
44. de Kunder, S.L., S.M.J. van Kuijk, K. Rijkers, et al. (2017) *Transforaminal lumbar interbody fusion (TLIF) versus posterior lumbar interbody fusion (PLIF) in lumbar spondylolisthesis: a systematic review and meta-analysis.* Spine J, **online**.
45. Carreon, L.Y., S.D. Glassman, Z. Ghogawala, et al. (2016) *Modeled cost-effectiveness of transforaminal lumbar interbody fusion compared with posterolateral fusion for spondylolisthesis using N(2)QOD data.* J Neurosurg Spine, **24**:916-21.
46. Soegaard, R., C.E. Bunger, T. Christiansen, et al. (2007) *Circumferential fusion is dominant over posterolateral fusion in a long-term perspective: cost-utility evaluation of a randomized controlled trial in severe, chronic low back pain.* Spine (Phila Pa 1976), **32**:2405-14.
47. Videbaek, T.S., F.B. Christensen, R. Soegaard, et al. (2006) *Circumferential fusion improves outcome in comparison with instrumented posterolateral fusion: long-term results of a randomized clinical trial.* Spine (Phila Pa 1976), **31**:2875-80.
48. Habicht, A., *Medicinsk statistik*. 1. udgave ed. 2017, Kbh.: Munksgaard. 244.
49. Fairbank, J.C. and P.B. Pynsent (2000) *The Oswestry Disability Index.* Spine (Phila Pa 1976), **25**:2940-52; discussion 2952.
50. Lawlis, G.F., R. Cuencas, D. Selby, et al. (1989) *The development of the Dallas Pain Questionnaire. An assessment of the impact of spinal pain on behavior.* Spine (Phila Pa 1976), **14**:511-6.
51. Ware, J.E., Jr. (2000) *SF-36 health survey update.* Spine (Phila Pa 1976), **25**:3130-9.
52. Manniche, C., K. Asmussen, B. Lauritsen, et al. (1994) *Low Back Pain Rating scale: validation of a tool for assessment of low back pain.* Pain, **57**:317-26.
53. Birkmeyer, N.J., J.N. Weinstein, A.N. Tosteson, et al. (2002) *Design of the Spine Patient outcomes Research Trial (SPORT).* Spine (Phila Pa 1976), **27**:1361-72.
54. Christensen, F.B. (2004) *Lumbar spinal fusion. Outcome in relation to surgical methods, choice of implant and postoperative rehabilitation.* Acta Orthop Scand Suppl, **75**:2-43.
55. Korsgaard, M., F.B. Christensen, K. Thomsen, et al. (2002) *The influence of lumbar lordosis on spinal fusion and functional outcome after posterolateral spinal fusion with and without pedicle screw instrumentation.* J Spinal Disord Tech, **15**:187-92.
56. Andersen, T., F.B. Christensen and C. Bunger (2006) *Evaluation of a Dallas Pain Questionnaire classification in relation to outcome in lumbar spinal fusion.* Eur Spine J, **15**:1671-85.

57. Zanoli, G., B. Stromqvist and B. Jonsson (2001) *Visual analog scales for interpretation of back and leg pain intensity in patients operated for degenerative lumbar spine disorders.* Spine (Phila Pa 1976), **26**:2375-80.
58. Hagg, O., P. Fritzell, A. Nordwall, et al. (2003) *The clinical importance of changes in outcome scores after treatment for chronic low back pain.* Eur Spine J, **12**:12-20.
59. Bjørner, J.B., *Dansk manual til SF-36 : et spørgeskema om helbredsstatus*. 1997, S.l.: Lif. 143 s. + 1 diskette i lomme.
60. Bjorner, J.B., K. Thunedborg, T.S. Kristensen, et al. (1998) *The Danish SF-36 Health Survey: translation and preliminary validity studies.* J Clin Epidemiol, **51**:991-9.
61. Brazier, J., J. Roberts and M. Deverill (2002) *The estimation of a preference-based measure of health from the SF-36.* J Health Econ, **21**:271-92.
62. Uden, A., M. Astrom and H. Bergenudd (1988) *Pain drawings in chronic back pain.* Spine (Phila Pa 1976), **13**:389-92.
63. Margolis, R.B., R.C. Tait and S.J. Krause (1986) *A rating system for use with patient pain drawings.* Pain, **24**:57-65.
64. Sahl Andersen, J., N. De Fine Olivarius and A. Krasnik (2011) *The Danish National Health Service Register.* Scandinavian Journal of Public Health, **39**:34-37.
65. Christensen, A., K. Hoy, C. Bunger, et al. (2014) *Transforaminal lumbar interbody fusion vs. posterolateral instrumented fusion: cost-utility evaluation along side an RCT with a 2-year follow-up.* Eur Spine J, **23**:1137-43.
66. Institut., S.S. (2012), *Tariffs in 2012*.
67. Institut, S.S. (2013), *The National Health Insurance Service Register 2013*: Available from: http://www.ssi.dk/Sundhedsdataogit/Registre/Sygesikringsregister.aspx.
68. Statistics Denmark. Disposable income by region, g., age and income range 2013 [cited 2013 May 24th]. Available from: http://www.statistikbanken.dk/statbank5a/selectvarval/define.asp?PLanguage=0&subword=tabsel&MainTable=INDKP6&PXSId=146749&tablestyle=&ST=SD&buttons=0 (2013).
69. Fritzell, P., O. Hagg, P. Wessberg, et al. (2001) *2001 Volvo Award Winner in Clinical Studies: Lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group.* Spine (Phila Pa 1976), **26**:2521-32; discussion 2532-4.
70. Hoy, K., K. Truong, T. Andersen, et al. (2017) *Addition of TLIF does not improve outcome over standard posterior instrumented fusion. 5-10 years long-term Follow-up: results from a RCT.* Eur Spine J, **26**:658-665.
71. Etemadifar, M.R., A. Hadi and M.F. Masouleh (2016) *Posterolateral instrumented fusion with and without transforaminal lumbar interbody fusion for the treatment of adult isthmic spondylolisthesis: A randomized clinical trial with 2-year follow-up.* J Craniovertebr Junction Spine, **7**:43-9.
72. Zhang, B.F., C.Y. Ge, B.L. Zheng, et al. (2016) *Transforaminal lumbar interbody fusion versus posterolateral fusion in degenerative lumbar spondylosis: A meta-analysis.* Medicine (Baltimore), **95**:e4995.
73. Grob, D., V. Bartanusz, D. Jeszenszky, et al. (2009) *A prospective, cohort study comparing translaminar screw fixation with transforaminal lumbar interbody fusion and pedicle screw fixation for fusion of the degenerative lumbar spine.* J Bone Joint Surg Br, **91**:1347-53.
74. Pannell, W.C., D.D. Savin, T.P. Scott, et al. (2015) *Trends in the surgical treatment of lumbar spine disease in the United States.* Spine J, **15**:1719-27.
75. Videbaek, T.S. (2010), *Long-term effect of anterior colunm support for mechanocal low back pain.*, in *Faculty of Health Sciences, Aarhus University, PhD Thesis*, Aarhus University: Aarhus Denmark.

76. Glassman, S.D., L.Y. Carreon, Z. Ghogawala, et al. (2016) *Benefit of Transforaminal Lumbar Interbody Fusion vs Posterolateral Spinal Fusion in Lumbar Spine Disorders: A Propensity-Matched Analysis From the National Neurosurgical Quality and Outcomes Database Registry.* Neurosurgery, **79**:397-405.
77. Epstein, N.E. (2016) *More nerve root injuries occur with minimally invasive lumbar surgery, especially extreme lateral interbody fusion: A review.* Surg Neurol Int, **7**:S83-95.
78. Epstein, N.E. (2016) *More nerve root injuries occur with minimally invasive lumbar surgery: Let's tell someone.* Surg Neurol Int, **7**:S96-S101.
79. Villavicencio, A.T., S. Burneikiene, K.R. Bulsara, et al. (2006) *Perioperative complications in transforaminal lumbar interbody fusion versus anterior-posterior reconstruction for lumbar disc degeneration and instability.* J Spinal Disord Tech, **19**:92-7.
80. Ohnmeiss, D.D., H. Vanharanta and J. Ekholm (1999) *Relationship of pain drawings to invasive tests assessing intervertebral disc pathology.* Eur Spine J, **8**:126-31.
81. Ohnmeiss, D.D., H. Vanharanta and J. Ekholm (1999) *Relation between pain location and disc pathology: a study of pain drawings and CT/discography.* Clin J Pain, **15**:210-7.
82. Ohnmeiss, D.D. (2000) *Repeatability of pain drawings in a low back pain population.* Spine (Phila Pa 1976), **25**:980-8.
83. Andersen, T., F.B. Christensen, K.W. Hoy, et al. (2010) *The predictive value of pain drawings in lumbar spinal fusion surgery.* Spine J, **10**:372-9.
84. Desai, A., P.A. Ball, K. Bekelis, et al. (2012) *Surgery for lumbar degenerative spondylolisthesis in Spine Patient Outcomes Research Trial: does incidental durotomy affect outcome?* Spine (Phila Pa 1976), **37**:406-13.
85. Hee, H.T., F.P. Castro, Jr., M.E. Majd, et al. (2001) *Anterior/posterior lumbar fusion versus transforaminal lumbar interbody fusion: analysis of complications and predictive factors.* J Spinal Disord, **14**:533-40.
86. Cole, C.D., T.D. McCall, M.H. Schmidt, et al. (2009) *Comparison of low back fusion techniques: transforaminal lumbar interbody fusion (TLIF) or posterior lumbar interbody fusion (PLIF) approaches.* Curr Rev Musculoskelet Med, **2**:118-26.
87. Teng, I., J. Han, K. Phan, et al. (2017) *A meta-analysis comparing ALIF, PLIF, TLIF and LLIF.* J Clin Neurosci, **44**:11-17.
88. Videbaek, T.S., N. Egund, F.B. Christensen, et al. (2010) *Adjacent segment degeneration after lumbar spinal fusion: the impact of anterior column support: a randomized clinical trial with an eight- to thirteen-year magnetic resonance imaging follow-up.* Spine (Phila Pa 1976), **35**:1955-64.
89. Mannion, A.F., G. Leivseth, J.I. Brox, et al. (2014) *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 (Phila Pa 1976), **39**:1373-83.
90. Soegaard, R. and F.B. Christensen (2006) *Health economic evaluation in lumbar spinal fusion: a systematic literature review anno 2005.* Eur Spine J, **15**:1165-73.
91. Bydon, M., M. Macki, N.B. Abt, et al. (2015) *The cost-effectiveness of interbody fusions versus posterolateral fusions in 137 patients with lumbar spondylolisthesis.* Spine J, **15**:492-8.
92. Rodriguez-Vela, J., A. Lobo-Escolar, E. Joven, et al. (2013) *Clinical outcomes of minimally invasive versus open approach for one-level transforaminal lumbar interbody fusion at the 3- to 4-year follow-up.* Eur Spine J, **22**:2857-63.
93. Swan, J., E. Hurwitz, F. Malek, et al. (2006) *Surgical treatment for unstable low-grade isthmic spondylolisthesis in adults: a prospective controlled study of posterior instrumented fusion compared with combined anterior-posterior fusion.* Spine J, **6**:606-14.

94. Villavicencio, A.T., S. Burneikiene, C.M. Roeca, et al. (2010) *Minimally invasive versus open transforaminal lumbar interbody fusion.* Surg Neurol Int, **1**:12.
95. Pelton, M.A., F.M. Phillips and K. Singh (2012) *A comparison of perioperative costs and outcomes in patients with and without workers' compensation claims treated with minimally invasive or open transforaminal lumbar interbody fusion.* Spine (Phila Pa 1976), **37**:1914-9.
96. Fritzell, P. (2005) *Fusion as treatment for chronic low back pain--existing evidence, the scientific frontier and research strategies.* Eur Spine J, **14**:519-20.
97. Juul, S.f., *Epidemiologi og evidens*. 3. udgave ed. 2017, Kbh.: Munksgaard. 328.
98. Høy, K., B. Grycel, T. Andersen, et al. (2013) *Tlif surgery results in slightly higher risk of neurogenic leg pain 2 years after surgery compared to standard instrumented posterolateral fusion. Results from a randomized clinical trial.* European Spine Journal, **22**:S717-S718.
99. Lauridsen, H.H., J. Hartvigsen, C. Manniche, et al. (2006) *Danish version of the Oswestry Disability Index for patients with low back pain. Part 1: Cross-cultural adaptation, reliability and validity in two different populations.* European Spine Journal, **15**:1705-1716.
100. Lauridsen, H.H., J. Hartvigsen, C. Manniche, et al. (2006) *Danish version of the Oswestry disability index for patients with low back pain. Part 2: Sensitivity, specificity and clinically significant improvement in two low back pain populations.* Eur Spine J, **15**:1717-28.
101. Bombardier, C. (2000) *Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations.* Spine (Phila Pa 1976), **25**:3100-3.
102. Andersen, T., F.B. Christensen, E.S. Hansen, et al. (2003) *Pain 5 years after instrumented and non-instrumented posterolateral lumbar spinal fusion.* Eur Spine J, **12**:393-9.
103. Andersen, T., T.S. Videbaek, E.S. Hansen, et al. (2008) *The positive effect of posterolateral lumbar spinal fusion is preserved at long-term follow-up: a RCT with 11-13 year follow-up.* Eur Spine J, **17**:272-80.
104. Marty, M., D. Courvoisier, V. Foltz, et al. (2016) *How much does the Dallas Pain Questionnaire score have to improve to indicate that patients with chronic low back pain feel better or well?* Eur Spine J, **25**:304-9.
105. Marty, M., F. Blotman, B. Avouac, et al. (1998) *Validation of the French version of the Dallas Pain Questionnaire in chronic low back pain patients.* Rev Rhum Engl Ed, **65**:126-34.
106. Ekman, P., H. Moller, A. Shalabi, et al. (2009) *A prospective randomised study on the long-term effect of lumbar fusion on adjacent disc degeneration.* Eur Spine J, **18**:1175-86.
107. Levison, D.R. (2013) *Spinal Devices Supplied by Physician-Owned Distributors: Overview of Prevalence and Use (OEI-01-11-00660).*
108. Whoriskey, P. (November 13, 2013) *Rise in spinal fusion surgeries may be driven partly by financial incentives, study says.* The Washington Post.
109. Mannion, A.F., J.I. Brox and J.C. Fairbank (2016) *Consensus at last! Long-term results of all randomized controlled trials show that fusion is no better than non-operative care in improving pain and disability in chronic low back pain.* Spine J, **16**:588-90.
110. Price, J.P., J.M. Dawson, J.D. Schwender, et al. (2017) *Clinical and Radiologic Comparison of Minimally Invasive Surgery With Traditional Open Transforaminal Lumbar Interbody Fusion: A Review of 452 Patients From a Single Center.* Clin Spine Surg.
111. Hackenberg, L., H. Halm, V. Bullmann, et al. (2005) *Transforaminal lumbar interbody fusion: a safe technique with satisfactory three to five year results.* Eur Spine J, **14**:551-8.
112. Fujimori, T., H. Le, W.W. Schairer, et al. (2015) *Does Transforaminal Lumbar Interbody Fusion Have Advantages over Posterolateral Lumbar Fusion for Degenerative Spondylolisthesis?* Global Spine J, **5**:102-9.

113. Ghasemi, A.A. (2016) *Transforaminal lumbar interbody fusion versus instrumented posterolateral fusion In degenerative spondylolisthesis: An attempt to evaluate the superiority of one method over the other.* Clin Neurol Neurosurg, **150**:1-5.
114. Forsth, P., G. Olafsson, T. Carlsson, et al. (2016) *A Randomized, Controlled Trial of Fusion Surgery for Lumbar Spinal Stenosis.* N Engl J Med, **374**:1413-23.
115. Mehta, V.A., M.J. McGirt, G.L. Garces Ambrossi, et al. (2011) *Trans-foraminal versus posterior lumbar interbody fusion: comparison of surgical morbidity.* Neurol Res, **33**:38-42.
116. Moher, D., S. Hopewell, K.F. Schulz, et al. (2010) *CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials.* BMJ, **340**.
117. Kirkaldy-Willis, W.H., J.H. Wedge, K. Yong-Hing, et al. (1982) *Lumbar spinal nerve lateral entrapment.* Clin Orthop Relat Res, 171-8.

Appendix articles:

Study 1:

Transforaminal lumbar interbody fusion (TLIF) versus posterolateral instrumented fusion (PLF) in degenerative lumbar disorders: a randomized clinical trial with 2-year follow-up

Eur Spine J (2013) 22:2022–2029
DOI 10.1007/s00586-013-2760-2

ORIGINAL ARTICLE

Transforaminal lumbar interbody fusion (TLIF) versus posterolateral instrumented fusion (PLF) in degenerative lumbar disorders: a randomized clinical trial with 2-year follow-up

Kristian Høy · Cody Bünger · Bent Niederman · Peter Helmig · Ebbe Stender Hansen · HaiSheng Li · Thomas Andersen

Received: 22 July 2012 / Revised: 2 February 2013 / Accepted: 18 March 2013 / Published online: 13 April 2013

Abstract
Purpose The aim of the present study was to analyze outcome, with respect to functional disability, pain, fusion rate, and complications of patients treated with transforaminal lumbar interbody fusion (TLIF) in compared to instrumented poserolateral fusion (PLF) alone, in low back pain. Spinal fusion has become a major procedure worldwide. However, conflicting results exist. Theoretical circumferential fusion could improve functional outcome. However, the theoretical advantages lack scientific documentation.
Methods Prospective randomized clinical study with a 2-year follow-up period. From November 2003 to November 2008 100 patients with severe low back pain and radicular pain were randomly selected for either posterolateral lumbar fusion [titanium TSRH (Medtronic)] or transforaminal lumbar interbody fusion [titanium TSRH (Medtronic)] with anterior intervertebral support by tantalum cage (Implex/Zimmer). The primary outcome scores were obtained using Dallas Pain Questionnaire (DPQ), Oswestry disability Index, SF-36, and low back pain Rating Scale. All measures assessed the endpoints at 2-year follow-up after surgery.
Results The overall follow-up rate was 94 %. Sex ratio was 40/58. 51 patients had TLIF, 47 PLF. Mean age 49(TLIF)/45(PLF). No statistic difference in outcome between groups could be detected concerning daily activity, work leisure, anxiety/depression or social interest. We found no statistic difference concerning back pain or leg pain. In both the TLIF and the PLF groups the patients had significant improvement in functional outcome, back pain, and leg pain compared to preoperatively. Operation time and blood loss in the TLIF group were significantly higher than in the PLF group ($p < 0.001$). No statistic difference in fusion rates was detected.
Conclusions Transforaminal interbody fusion did not improve functional outcome in patients compared to posterolateral fusion. Both groups improved significantly in all categories compared to preoperatively. Operation time and blood loss were significantly higher in the TLIF group.

Keywords Prospective · RCT · Lumbar interbody fusion

Introduction

Spinal fusion has for many years been the treatment of choice to treat low back pain generated from disc degeneration, failed disc surgery, spondylosis, spondylolisthesis, resistant to conservative therapy, and in combination with decompressions performed due to spinal stenosis. Different methods have evolved over the last many years. At the present time, there is no clear cut scientific evidence of which surgical strategy is the best [9]. There have been randomized controlled trials showing no significant improvement in short time outcomes in instrumented fusion compared to uninstrumented fusion [8, 16, 19]. Nevertheless treatment strategies have moved towards global fusion based on the theoretical point of view that restoration of lordosis, sagittal balance, and neuroforaminal decompression due to restoration of the disc height would result in better functional outcomes. However, this theory has been difficult to validate scientifically. During the late 90s and in the beginning of the new century, anterior lumbar interbody

K. Høy (✉) · C. Bünger · B. Niederman · P. Helmig · E. S. Hansen · H. Li · T. Andersen
Spine Section, Departments of Orthopedics E, Aarhus University Hospital, Aarhus University, Nørrebrogade 44, Aarhus C 8000, Denmark
e-mail: wintherhoey@dadlnet.dk; krishoey@rm.dk

fusion (ALIF) was used to minimize pain in patients suffering from chronic low back pain (CLBP). However, only one randomized clinical trial (RCT) with long time follow-up has been able to show benefits of this procedure [20]. In comparison, the Swedish Lumbar Spine Study reported significantly higher complication rates in the interbody fusion group compared to the instrumented and non-instrumented posterolateral fusion groups and no difference in functional outcome between groups [7]. The need of anterior support has also been questioned by Ekman and colleagues [4], comparing posterior lumbar interbody fusion (PLIF) to posterolateral lumbar fusion (PLF) in adult isthmic spondylolisthesis: they did not find any difference in patient reported outcome. Recently, Neumann et al. [17] presented data from an RCT comparing transforaminal lumbar interbody fusion (TLIF) to uninstrumented PLF in degenerative disc disease (DDD), where outcome was significantly in favour of TLIF in visual analogue scale (VAS, pain) and disability Rating Index, but with no significant difference in Oswestry disability Index (ODI). Our group has earlier been able to show benefits of 360° fusion utilizing anterior interbody fusion [3]. Transforaminal lumbar interbody fusion (TLIF) should theoretically offer the same benefits of circumferential fusion, but with less morbidity due to the lesser surgical approach. But as there is conflicting evidence, the aim of the present study was designed to test whether a TLIF procedure resulted in better functional outcome than PLF instrumented posterolateral fusion.

Materials and methods

From November 2003 through November 2008, a total of 100 patients (mean age 49.8 years; TLIF group 50.3 years, PLF group 49.3 years) were included in the present single-centre prospective RCT. The study protocol was approved by the regional ethical committee (No: 20030172). All patients suffered from severe CLBP and/or leg pain, static or dynamic, resulting from localized lumbar or lumbosacral segmental instability, spinal stenosis at levels L2–S1 or caused by isthmic spondylolisthesis (grade 1 and 2). Baseline characteristics concerning demographic and clinical data are presented in Table 1. Of the 14 patients in the TLIF group, who had undergone previous spine surgery, ten had had a prior discectomy and four a decompressive procedure; in the control group there were seven discectomies, three fusions and three decompressive procedures. In the TLIF group 21 patients (41 %) used opioids and six patients (12 %) used anti-depressants before surgery; in the control group it was 21 (43 %) and 8 (16 %), respectively (differences non-significant).

Table 1 Patient characteristics according to treatment group

	TLIF	Control	*p* value
Sex (female/male)	27/24	32/17	0.209
Age at surgery (years)	51 (30–63)	49 (25–70)	0.269
Diagnosis			0.335
Spondylolisthesis	10	17	
DDD	20	17	
Spinal stenosis	9	8	
Failed back surgery	12	7	
Pain duration			0.283
<1 year	8	3	
1–2 years	14	17	
>2 years	29	29	
Work status			0.740
Working	22	16	
Without work	2	2	
Sick-leave	14	15	
Retired	13	16	
Smoking	24 (44 %)	21 (46 %)	0.904
Previous spine surgery	14 (27 %)	13 (27 %)	0.917
Ongoing case[a]	16 (31 %)	7 (14 %)	0.042
Spondylolisthesis	1	2	
DDD	6	1	
Spinal stenosis	4	0	
Failed back surgery	5	4	
Operated level(s)			0.479
1 level	36	29	
2 levels	14	19	
3 levels	1	1	
Additional neural decompression			<0.001
None	0	18	
Laminotomy	33	7	
Spondylolisthesis	3	0	
DDD	15	1	
Spinal stenosis	4	2	
Failed back surgery	11	4	
Laminectomy	18	24	
Spondylolisthesis	7	14	
DDD	5	4	
Spinal stenosis	5	5	
Failed back surgery	1	1	
Operation time (min)	228 (135–450)	171 (100–300)	<0.001
Blood loss (ml)	775 (150–5,000)	443 (50–1,500)	0.001
Hospitalisation (days)	9.8 (5–22)	9.3 (6–16)	0.405

Values are mean (range) or number (%)

[a] Ongoing case means: insurance, compensatory case, negligence case, etc

Before surgery, the indication for fusion was determined on the basis of anamnesis, several clinical and neurological examinations, and MRI. Exclusion criteria comprised age younger than 25 years, spondylolisthesis grade III and IV, osteoporosis diagnosed via radiography and bone mineral density testing, severe cardiac or vascular disease, brain disorders, kidney problems, former or actual malignancy, use of medicine reducing bone metabolism, dementia or abnormal psychological behaviour, and language problems.

Randomization was done using sealed envelopes with a 20-number-per block randomization. The envelopes were consecutively numbered, thereby assigning a number to each patient in the study. The type of operation remained unknown to both the patient and the surgeon until the patient's written consent was obtained. By this procedure, 49 patients were allocated to posterolateral fusion with titanium TSRH (Medtronic) pedicle instrumentation (PLF group) (Fig. 1) and 51 patients to transforaminal fusion in the form of a tantalum cage (Implex/Zimmer) placed using an approach lateral to the facet joint (TLIF group) (Fig. 2). The anterior interbody fusion device was supported by a posterolateral fusion using pedicle screws (titanium TSRH, Medtronic).

During surgery the patients were placed in prone position. We used controlled hypotensive anaesthesia. The patients first underwent insertion of pedicle screws by a midline subperiosteal approach. When indicated, hemilaminectomy or laminectomy for neural decompression was performed. In case the patients were randomized to the transforaminal procedure, the facet joint of the intended

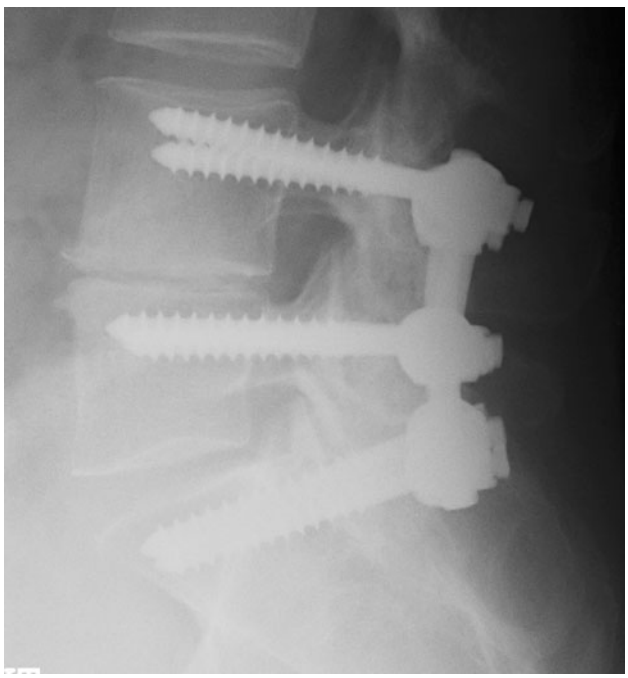


Fig. 1 X-ray showing the instrumentation in the control group

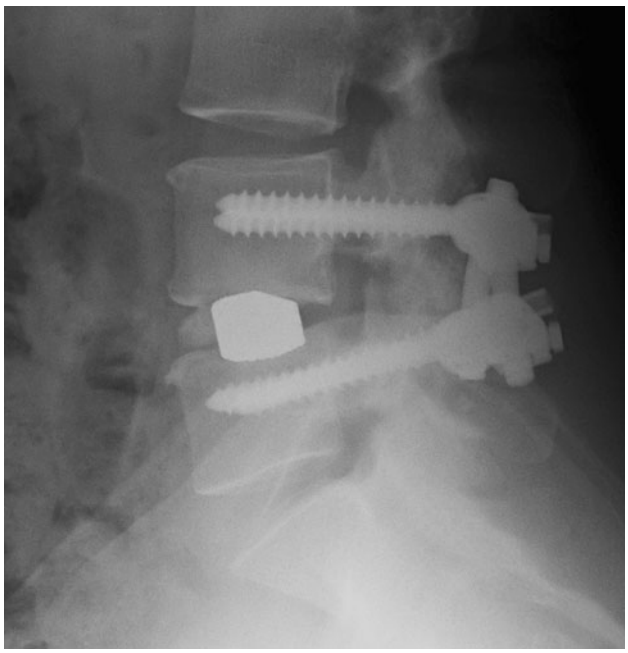


Fig. 2 X-ray showing the instrumentation in the TLIF group

levels was indentified and the inferior and superior facets were resected to gain access to the disc space, and by that procedure an indirectly neurolysis or decompression of the nerve was performed. The pedicle screws were used to distract. The upper nerve was indentified and protected. The tantalum cage was placed after cleaning the disc space. Compression over the disc space was done after placement of the cage in order to create lordosis [11]. Cancellous bone from bone bank femoral heads were used as bone graft and placed on the transverse process of the vertebraes fused. A careful preparation of the posterolateral region was performed before positioning of the graft. Before that the decompressed neural structures were covered with a gelfoam (Spongostan) to avoid damage.

Overall, a 2-year follow-up rate of 94 % was achieved, with follow-up evaluation for 47 patients in the PLF group and 47 patients in the TLIF group. One patient from the PLF group had moved to Greenland and did not show up for follow-up. One patient from the TLIF group had died for reasons unrelated to the surgery. One patient from the PLF group and one patient from the TLIF group did not show up to the follow-up due to disappointment with the effect of the procedure. Two patients from the TLIF group did not answer follow-up.

Functional outcome was assessed by means of the Dallas Pain Questionnaire (DPQ), the pain index from the low back pain rating Scale (LBPRS), ODI, and SF-36. The questionnaires were completed by the patients, independently of the surgeon, before the operation and at the 1 and 2 years follow-up assessments. The DPQ is validated and it

assesses the functional impact of chronic spinal pain in four categories: daily activity, work-leisure activity, anxiety-depression, and social concerns. A high score indicates great functional impact [13]. Back and leg pain was assessed by the LBPRS pain index. It is a validated index scale that includes measurements of pain intensity ranging from 0 to 10, in which 10 represents the worst possible pain. They are 11 point (0–10) numerical rating scales assessing both back and leg pain in three ways: worst pain within in the last 14 days, average pain within the last 14 days, and actual pain level at the time of completing the questionnaire. The scores for leg pain and back pain are summed giving a pain index ranging from 0 to 60 [14]. The ODI is a condition specific outcome measure for spinal disorders, yielding an index score which ranges from 0 to 100, with a high percentage reflecting a high degree of disability [5]. The SF-36 is a generic health survey measure. It yields a profile of scores in eight scales covering different physical and mental components of health. The score in each scale ranges from 0 (poorest health) to 100 (best health). In addition two summary measures are produced: a physical component summary (PCS) and a mental component summary (MCS). The two summary measures are calculated so that the value of 50 is equal to the US population mean [22]. Before surgery and at the follow-up each patient received a questionnaire concerning work status and employment. At follow-up assessments patients were also asked the question "would you undergo the same treatment again now you know the result?", the answer to which served as a global outcome parameter.

All analyses comparing the two intervention groups were done using the intention to treat principle. Comparison between groups were done using non-parametric tests (Mann–Whitney rank-sum test for unpaired data, Wilcoxon signed rank test for paired data, Chi-square statistics or the Kruskal–Wallis test for equality at groups, with correction for ties) depending on the nature of the data. Significance level was 5 % using two-tailed testing. Intercooled Stata version 12 for Windows was the software used for statistical analysis. Sample size of the study was calculated using a significance level of 0.05 and a power of 0.80. Based on earlier studies, the standard deviation of the DPQ daily activity score was set at 25 points. A 15-point difference in this category was considered clinically relevant. To fulfill these criteria, the study would need 44 patients in each group. To allow for dropouts, the study was designed with 50 patients in each group.

Results

We could not observe any difference in favour of the TLIF procedure in any of the outcome parameters measured at any time point (Figs. 1, 2). Both groups improved significantly compared to preoperatively, but without any difference between the groups (Fig. 1). There was, however, an insignificant trend towards the TLIF group developing more leg pain during follow-up compared to the controls (Table 2). Subgroup analysis based on diagnosis could not reveal any substantial benefits of the TLIF procedure either (Table 3). Number of patients with ongoing case was the only variable to differ between the two randomization groups (Table 1). After stratifying for ongoing compensatory case still no difference between the two procedures could be observed (data not shown).

Data on work status at 2-year follow-up were available in 96 patients. In the TLIF group 23 patients were working, one was without work, two were on sick-leave and 22 were retired. In the control group the numbers were 23 working, 2 on sick-leave and 27 retired ($p = 0.60$). In the TLIF group 16 patients used opioids and three patients antidepressants at the 2-year follow-up compared to 14 and 6 patients, respectively, in the control group (not significant). Fusion rate at 2 years was 94 % (44 of 47 patients with available radiographs) in the TLIF group compared to 88 % (42 of 48 patients with available radiographs) in the control group ($p = 0.31$). 91 patients, 46 in the TLIF group and 45 in the control group had answered the question "Knowing the result would then undergo the procedure again?" at 2 years. Thirty-three patients in each group

Table 2 Scores at all time points for the subcategories of the LBPRS

	Preoperative			1-year follow-up			2-year follow-up		
	TLIF	Control	*p* value	TLIF	Control	*p* value	TLIF	Control	*p* value
Back pain right now	6.1 (0.3)	5.8 (0.3)	0.535	3.6 (0.5)	3.4 (0.4)	0.981	3.5 (0.4)	3.6 (0.5)	1.000
Worst back pain within last 14 days	7.8 (0.3)	8.1 (0.3)	0.707	4.9 (0.5)	4.6 (0.4)	0.659	5.3 (0.5)	5.1 (0.5)	0.795
Average back pain within last 14 days	6.1 (0.4)	6.1 (0.3)	0.894	3.8 (0.4)	3.8 (0.4)	0.883	4.1 (0.4)	4.0 (0.5)	0.774
Leg pain right now	5.1 (0.4)	4.3 (0.4)	0.215	3.0 (0.5)	2.4 (0.4)	0.542	3.2 (0.4)	2.5 (0.4)	0.265
Worst leg pain within last 14 days	6.2 (0.4)	6.5 (0.4)	0.657	4.0 (0.6)	3.2 (0.5)	0.470	4.4 (0.5)	3.4 (0.5)	0.137
Average leg pain within last 14 days	5.1 (0.4)	5.0 (0.4)	0.888	3.2 (0.5)	2.7 (0.4)	0.589	3.5 (0.4)	2.9 (0.5)	0.214

Values are mean (standard error of mean)

Table 3 Outcome scores at 2-year follow-up in the two treatment groups stratified according to surgical diagnosis

	TLIF	Control	*p* value
Spondylolisthesis			
DPQ daily activity	40.1 (11.1)	41.8 (9.1)	0.799
DPQ work/leisure activity	31.1 (10.4)	39.6 (11.5)	0.562
DPQ anxiety/depression	21.9 (10.7)	18.2 (4.8)	0.808
DPQ social interest	15.6 (5.8)	21.1 (7.2)	0.974
LBPRS pain index	19.6 (2.7)	17.1 (3.2)	0.503
ODI	25.8 (6.9)	26.7 (4.5)	0.698
SF-36 BP	50.8 (7.0)	57.5 (7.1)	0.509
SF-36 PF	60.8 (8.3)	64.3 (6.6)	0.976
SF-36 PCS	40.0 (3.2)	41.2 (3.5)	0.659
DDD			
DPQ daily activity	45.5 (7.4)	30.4 (7.3)	0.145
DPQ work/leisure activity	48.2 (9.2)	32.0 (7.6)	0.155
DPQ anxiety/depression	29.2 (7.2)	24.3 (6.7)	0.645
DPQ social interest	25.8 (7.5)	20.0 (5.6)	0.901
LBPRS pain index	26.3 (4.2)	25.0 (5.1)	0.600
ODI	32.4 (4.5)	21.8 (4.0)	0.144
SF-36 BP	48.7 (6.9)	51.9 (8.0)	0.701
SF-36 PF	63.1 (5.6)	70.2 (6.2)	0.431
SF-36 PCS	37.2 (2.4)	42.3 (2.8)	0.184
Spinal stenosis			
DPQ daily activity	36.4 (9.9)	38.1 (9.8)	0.954
DPQ work/leisure activity	37.5 (12.5)	35.0 (13.0)	0.942
DPQ anxiety/depression	15.6 (7.9)	28.6 (9.0)	0.143
DPQ social interest	18.1 (6.0)	10.7 (7.1)	0.231
LBPRS pain index	23.6 (7.2)	18.1 (4.5)	0.728
ODI	25.0 (6.6)	27.3 (8.5)	1.000
SF-36 BP	52.6 (11.1)	47.4 (11.2)	0.728
SF-36 PF	56.4 (9.7)	62.1 (9.9)	0.560
SF-36 PCS	35.5 (5.4)	38.5 (4.3)	0.729
Failed back surgery			
DPQ daily activity	39.0 (10.3)	54.4 (12.0)	0.222
DPQ work/leisure activity	42.5 (12.6)	53.8 (16.6)	0.670
DPQ anxiety/depression	32.8 (14.1)	45.8 (13.9)	0.551
DPQ social interest	34.5 (13.5)	28.3 (9.7)	0.912
LBPRS pain index	24.0 (4.7)	25.8 (9.5)	0.957
ODI	34.4 (8.0)	35.7 (5.1)	0.625
SF-36 BP	48.5 (8.0)	53.0 (10.4)	0.404
SF-36 PF	61.6 (11.0)	50.7 (8.9)	0.463
SF-36 PCS	39.5 (4.1)	35.0 (4.4)	0.540

Values are mean (standard error of mean)

answered with a positive response (72 vs. 73 %, $p = 0.87$) (Figs. 3, 4).

Operation time and blood loss were significantly higher in the TLIF group, but did not result in longer length of hospitalization (Table 1). The complication rate in the TLIF group was 14 % (7 patients) as compared to 6 % (3 patients) in the control group ($p = 0.205$). In the TLIF group complications were one haematoma, two superficial infections, one nerve root lesion, two dural tears, and one intraoperatively developed pneumothorax, which was treated with drainage. In the control group, there were two haematomas and one dural tear. In two of the TLIF cases the disc space was too small for insertion of the cage and instead the disc space was filled with allograft bone. Also in one case in the TLIF group it was preoperatively decided to convert to an anterior procedure. Among the study patients nine patients (9 %) had undergone a second operation: three patients in PLF group had removal of the hardware due to loosening or failure. Four patients in the TLIF group had removal of the implant: two due to misplaced cages, one due to non-union, (this patient subsequently had an ALIF), and one due to deep infection. In addition, two patients in the TLIF group had secondary decompression at another level performed within 1 year.

Discussion

Transforaminal lumbar interbody fusion is today widely used in lumbar spinal fusion because of less violation to the spinal canal, compared to PLIF, and due to less time consumption and morbidity compared to ALIF, to achieve interbody fusion, which by many authors is considered to be the treatment of choice [2, 3, 15, 20, 23]. To our knowledge, this study is the first randomized prospective study to analyze a standardized instrumented spinal posterolateral fusion procedure with a TLIF procedure.

The present study was a single-centre study, which increases the possibility of a standardized patient selection and uniform surgical technique. This we believe improved the design and power of the present study. We could not demonstrate any superiority of the procedure with respect to function and back pain in a 2 years perspective. Neither could we demonstrate any improvement in leg pain in the TLIF group compared to the PLF group. On the contrary, there was a tendency towards more leg pain at 2-year follow-up in the TLIF group, questioning the concept of indirect foraminal decompression in the interbody fusion method or indicating that the procedure offers a risk of injury to the upper nerve root during access to and cleaning of the intervertebral disc space. Our study consisted of a mixed patient material with and without radicular pain, which could blur the results. Another major difference between the two methods in our study is that the TLIF procedure always is associated with a decompression at one side, irrespective of whether this is needed or not. In case of preoperative sciatica, the procedure was performed on the side with sciatica. The fact that the procedure always included decompression would also put the nerve root at

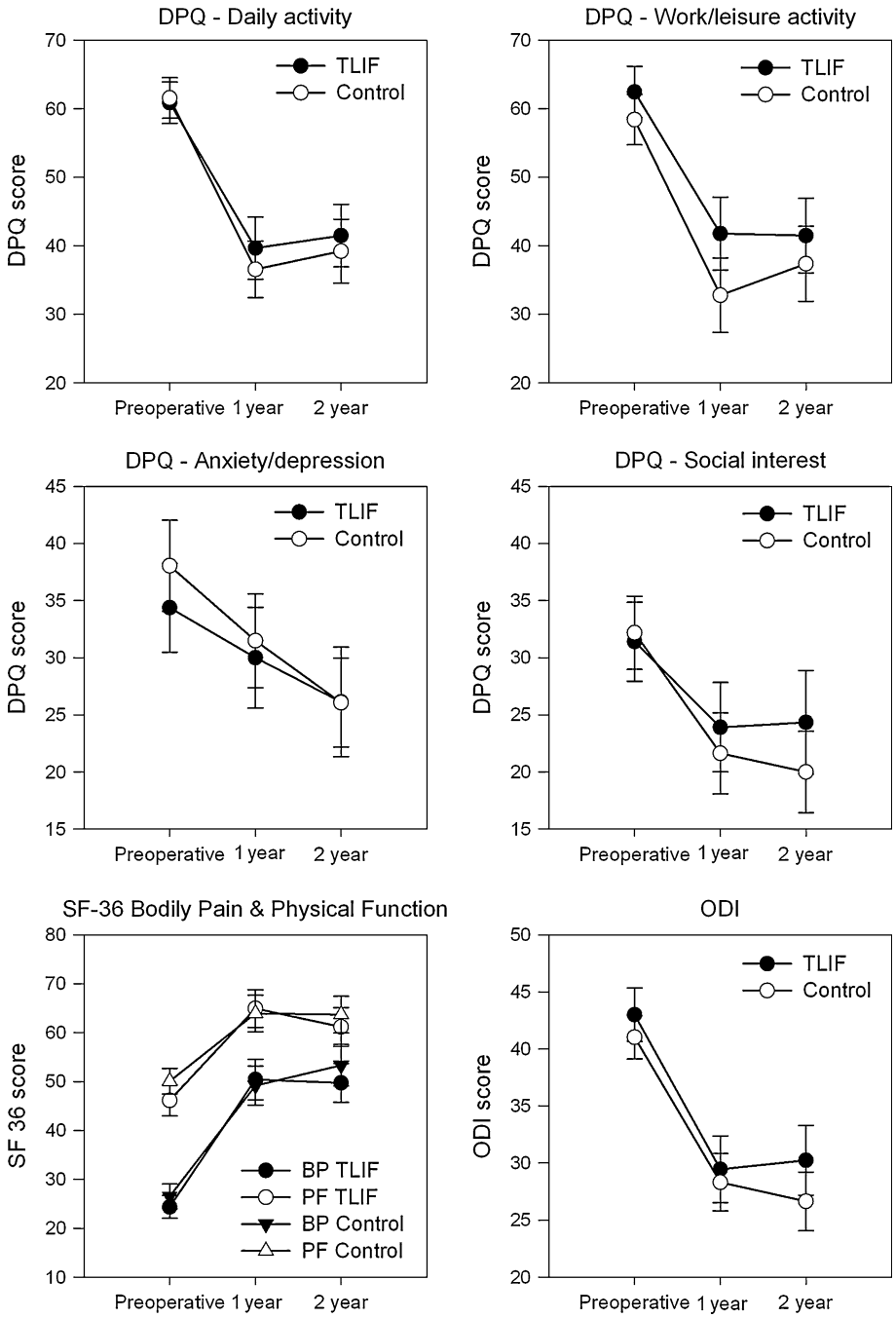


Fig. 3 Preoperative and 1 and 2 year follow-up scores of the four DPQ categories as well as ODI and the BP (bodily pain) and PF (physical function) subscales from the SF-36. There was no significant difference between the two groups at any time points ($p > 0.50$, except DPQ work/leisure activity $p > 0.29$). Both groups improved significantly from preoperatively to last follow-up ($p < 0.01$, except DPQ anxiety/depression in the TLIF group $p = 0.03$). Values are mean and errors are standard error of the mean

risk of later irritation due to the formation of scar tissue. This could explain why we observed slightly more leg pain in the TLIF group at 2-year follow-up, and it could be an argument in favour of ALIF surgery.

The only variable which differed between the two randomization groups was the number of patients with an ongoing compensatory case. As this has been shown to be a risk factor for poorer outcome it might hide a small effect in favour of the TLIF procedure. But we could not prove any difference between the two procedures in the patients without ongoing compensatory case, suggesting that our results are not skewed significantly by this.

Only one RCT study has shown evidence to support the benefits of the TLIF procedure in DDD [17]. They found significant improvement in the TLIF group in pain index and in disability Rating Index compared to their control group. However, difference in the widely used ODI was not significantly in favour of TLIF. Unfortunately, the data from this study have not been published in full yet. In the abstract they concluded that the results strongly suggested

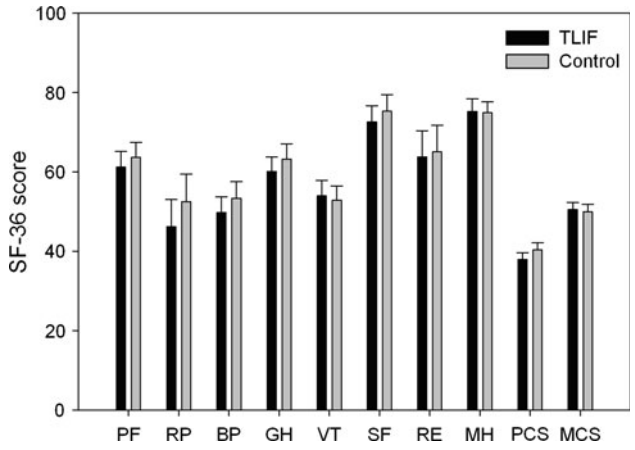


Fig. 4 Two year follow-up scores, according to randomization group, of the eight subscales and two summary measures of the SF-36. There was no significant difference between the two groups in any of the scores ($p > 0.30$). Values are mean and errors are standard error of mean

interbody fusion to be considered as the primary choice in the degenerative lumbar spine. This strong conclusion might not be valid considering our results. One explanation to the difference between these two studies could be the choice of control group. Their control group was uninstrumented fusion, a comparison that could be questioned, as it compares a very stable construct to an unstable construct. On the other hand, several studies have failed to show any difference between uninstrumented and instrumented posterolateral fusion, including the Swedish Spine Study, which also failed to show any effect of interbody fusion, performed as a PLIF or an ALIF procedure for patients with CLBP [6, 8, 19].

The positive effects of interbody fusion have been shown in a long-term follow-up study comparing 360° fusion, done as instrumented posterolateral fusion combined with an ALIF in a mixed population. In that study Videbaek et al. [20] showed significantly better functional outcome in the ALIF group, measured using DPQ, ODI, SF-36, leg and back pain NRS, compared to the control group comprising instrumented posterolateral fusion. However, in the 2-year follow-up of that study no functional outcome benefits could be proven, although significant better fusions rates and fewer reoperations were observed in the 360-group. Their long-term results are in contrast to the findings of Swan et al. [18] who were able to show a positive effect of anterior support in 6 months and at 1-year follow-up but not at 2-year follow-up in a mixed population. They concluded that the positive effects of global fusion diminished over time. Comparison of the three fusion techniques uninstrumented posterolateral, instrumented posterolateral and 360° fusion was also done in the SPORT study on degenerative spondylolisthesis [1]. It showed minor differences between the types of fusion, with a slight tendency towards 360° fusion performing better at 1 and 2 year follow-up, but with no difference at the 3 and 4 year follow-ups. This study was, however, performed in a population significantly older than the current study and the comparison between fusion techniques was done in a non-randomized fashion. Kim et al. [12] performed a randomized study comparing instrumented posterolateral fusion with a PLIF procedure, both with and without additional posterolateral bone graft in DDD, and found no significant differences with respect to pain and functional outcome. Neither could Grob et al. [10] demonstrate any advantages of a TLIF procedure compared to posterolateral fusion performed using translaminar facet screw fixation in a mixed population of DDD, degenerative spondylolisthesis and facet syndrome; this was, however, not a randomized study.

In our study, we did not observe a higher complication rate in the TLIF group compared to the PLF group, although it has been shown earlier that an increase in technicality leads to higher complications rates [4, 7, 21]. This could be due to the fact that our study was a single-centre study, possibly reflecting a more standardized patient selection and surgical technique, as compared to the multi-centre Swedish Spine Study, with 19 different orthopedic departments participating in the study, and with different frequencies of performing the procedure [7]. Neither could we observe any significant difference in re-operation as seen in the ALIF study by Christensen et al. [3]. On the contrary, the data from the SPORT study on degenerative spondylolisthesis did not show any difference in complication or repeated surgery rate up to 4 years, between the three fusion groups in that study [1].

These different studies have been performed in patients with varying diagnosis. This could explain some of the differences. The present study was performed in a group of patients with a mixed diagnosis and we could not demonstrate any difference between the two surgical techniques in any of the diagnostic groups, but one has to bear in mind that the study was not powered for subgroup analysis.

In conclusion, this study showed that TLIF demands more extensive resources than posterolateral fusion with pedicle screws, without significant improvement in functional outcome or fusion rate and with a tendency towards more leg pain.

Conflict of interest None.

References

1. Abdu WA, Lurie JD, Spratt KF et al (2009) Degenerative spondylolisthesis: does fusion method influence outcome? 4-year results of the spine patient outcomes research trial. Spine 34(21): 2351–2360

2. Brantigan JW, Neidre A, Toohey JS (2004) The lumbar I/F Cage for posterior lumbar interbody fusion with the variable screw placement system: 10-year results of a food and drug administration clinical trial. Spine J 4(6):681–688
3. Christensen FB, Hansen ES, Eiskjaer SP et al (2002) Circumferential lumbar spinal fusion with Brantigan cage versus posterolateral fusion With Titanium Cotrel-Dubousset instrumentation: a prospective, randomized Clinical study of 146 patients. Spine 27(23):2674–2683
4. Ekman P, Moller H, Tullberg T, Neumann P, Hedlund R (2007) Posterior lumbar interbody fusion versus posterolateral fusion in adult isthmic spondylolisthesis. Spine 32(20):2178–2183
5. Fairbank JC, Pynsent PB (2000) The Oswestry disability index. Spine 25(22):2940–2953
6. Fischgrund JS, Mackay M, Herkowitz HN et al (1997) Volvo award winner in clinical studies. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective, randomized study comparing decompressive laminectomy and arthrodesis with and without spinal instrumentation. Spine 22(24):2807–2812
7. Fritzell P, Hagg O, Nordwall A (2003) Complications in lumbar fusion surgery for chronic low back pain: comparison of three surgical techniques used in a prospective randomized study. A report from the Swedish lumbar spine study group. Eur Spine J 12(2):178–189
8. Fritzell P, Hagg O, Wessberg P, Nordwall A (2002) Chronic low back pain and fusion: a comparison of three surgical techniques: a prospective multicenter randomized study from the Swedish lumbar spine study group. Spine 27(11):1131–1141
9. Gibson JN, Waddell G (2005) Surgery for degenerative lumbar spondylosis: updated cochrane review. Spine 30(20):2312–2320
10. Grob D, Bartanusz V, Jeszenszky D et al (2009) A prospective, cohort study comparing translaminar screw fixation with transforaminal lumbar interbody fusion and pedicle screw fixation for fusion of the degenerative lumbar spine. J Bone Joint Surg Br 91(10):1347–1353
11. Harms JG, Jeszenszky D (1998) The unilateral transforaminal approach for posterior lumbar interbody fusion. Orthop Traumatol 6(2):88–89
12. Kim KT, Lee SH, Lee YH, Bae SC, Suk KS (2006) Clinical outcomes of three fusion methods through the posterior approach in the lumbar spine. Spine 31(12):1351–1357
13. Lawlis GF, Cuencas R, Selby D, McCoy CE (1989) The development of the Dallas Pain Questionnaire. An assessment of the impact of spinal pain on behavior. Spine 14(5):511–516
14. Manniche C, Asmussen K, Lauritsen B et al (1994) Low back pain rating scale: validation of a tool for assessment of low back pain. Pain 57(3):317–326
15. McAfee PC, Devine JG, Chaput CD et al (2005) The indications for interbody fusion cages in the treatment of spondylolisthesis: analysis of 120 cases. Spine (Phila Pa 1976) 30(6 Suppl):S60–S65
16. Moller H, Hedlund R (2000) Instrumented and noninstrumented posterolateral fusion in adult spondylolisthesis: a prospective randomized study: part 2. Spine 25(13):1716–1721
17. Neumann P, Jalalpour K, Johansson C, Hedlund R (2009) A RCT between transforaminal lumbar interbody fusion (TLIF) and posterolateral fusion in DDD and postdiscectomy syndrome. Eur Spine J 18(Supplement 4):S408
18. Swan J, Hurwitz E, Malek F et al (2006) Surgical treatment for unstable low-grade isthmic spondylolisthesis in adults: a prospective controlled study of posterior instrumented fusion compared with combined anterior-posterior fusion. Spine J 6(6):606–614
19. Thomsen K, Christensen FB, Eiskjaer SP et al (1997) Volvo award winner in clinical studies. The effect of pedicle screw instrumentation on functional outcome and fusion rates in posterolateral lumbar spinal fusion: a prospective, randomized clinical study. Spine 22(24):2813–2822
20. Videbaek TS, Christensen FB, Soegaard R et al (2006) Circumferential fusion improves outcome in comparison with instrumented posterolateral fusion: long-term results of a randomized clinical trial. Spine 31(25):2875–2880
21. Villavicencio AT, Burneikiene S, Bulsara KR, Thramann JJ (2006) Perioperative complications in transforaminal lumbar interbody fusion versus anterior-posterior reconstruction for lumbar disc degeneration and instability. J Spinal Disord Tech 19(2):92–97
22. Ware JE Jr (2000) SF-36 Health Survey Update. Spine 25(24):3130–3139
23. Zdeblick TA, Phillips FM (2003) Interbody cage devices. Spine 28(15):S2–S7

Study 2:

Addition of TLIF does not improve outcome over standard posterior instrumented fusion. 5–10 years long-term Follow-up: results from a RCT

Eur Spine J
DOI 10.1007/s00586-016-4592-3

ORIGINAL ARTICLE

Addition of TLIF does not improve outcome over standard posterior instrumented fusion. 5–10 years long-term Follow-up: results from a RCT

Kristian Høy[1] · Kamilla Truong[1] · Thomas Andersen[1] · Cody Bünger[1]

Received: 7 August 2015 / Revised: 27 April 2016 / Accepted: 27 April 2016

Abstract
Purpose The use of inter-body device in lumbar fusions has been difficult to validate, only few long-term RCT are available.
Methods Between 2003 and 2005, 100 patients entered a RCT between transforaminal lumbar inter-body fusion (TLIF) or posterolateral instrumented lumbar fusion (PLF). The patients suffered from LBP due to segmental instability, disc degeneration, former disc herniation, spondylolisthesis Meyerding grade <2. Functional outcome parameters as Dallas pain questionnaire (DPQ), SF-36, low back pain questionnaire (LBRS), Oswestry disability index (ODI) were registered prospectively, and after 5–10 years.
Results Follow-up reached 93 % of available, (94 %, 44 in the PLF's and 92 %, 44 in the TLIF group $p = 0.76$). Mean follow-up was 8.6 years (5–10 years). Mean age at follow-up was 59 years (34–76 years $p = 0.19$). Reoperation rate in a long-term perspective was equal among groups 14 %, each $p = 0.24$. Back pain was 3.8 (mean) (Scale 0–10), TLIF (3.65) PLF (3.97) $p = 0.62$, leg pain 2.68 (mean) (Scale 0–10) 2.90 (TLIF) and 2.48 (PLF) $p = 0.34$. No difference in functional outcome between groups $p = 0.93$. Overall, global satisfaction with the primary intervention at 8.6 year was 76 % (75 % TLIF and 77 % PLF) $p = 0.85$.
Conclusion In a long-term perspective, patients with TLIF's did not experience better outcome scores.

✉ Kristian Høy
krishoey@rm.dk; wintherhoey@dadlnet.dk

[1] Department of Orthopedics, Spine Section E, Aarhus University Hospital, Aarhus University, Nørrebrogade 44, 8000 Aarhus C, Denmark

Keywords Randomized clinical trial (RCT) · Transforaminal lumbar inter-body fusion (TLIF) · Posterolateral instrumented lumbar fusion (PLF) · Long-term · Functional outcome

Introduction

The literature reveals evidence that support surgical treatment over continuing conservative treatment in treatment of low back pain [1–5]. Spinal fusion has for many years been the treatment of choice to treat low back pain generated from disc degeneration, failed disc surgery, spondylosis, spondylolisthesis, resistant to conservative therapy, and in combination with decompressions performed due to spinal stenosis. Different methods have evolved over the last many years [6, 7]. Even though the number of spinal procedures due to low back pain has raised tremendously during the last 15 years, there is still no clear scientific evidence onto which surgical strategy is the best [8]. There have been randomized controlled trials showing no significant difference in improvement comparing instrumented fusion to un-instrumented fusion at short time follow-up [3, 5, 9]. Nevertheless, treatment strategies have moved towards global fusion based on the theoretical point of view that restoration of lordosis, sagittal balance, and decompression of the neuroforamina due to restoration of the disc height would result in better functional outcomes. Also, the removal of the supposedly pain generating degenerated disc has been claimed in favor of anterior support [10, 11]. However, this theory has been difficult to validate scientifically and only a few long-term randomized studies are available [5, 12, 13]. In the late 90s and in the beginning of the new century, ALIF (anterior lumbar inter-body fusion) was used to minimize pain in

Published online: 07 May 2016

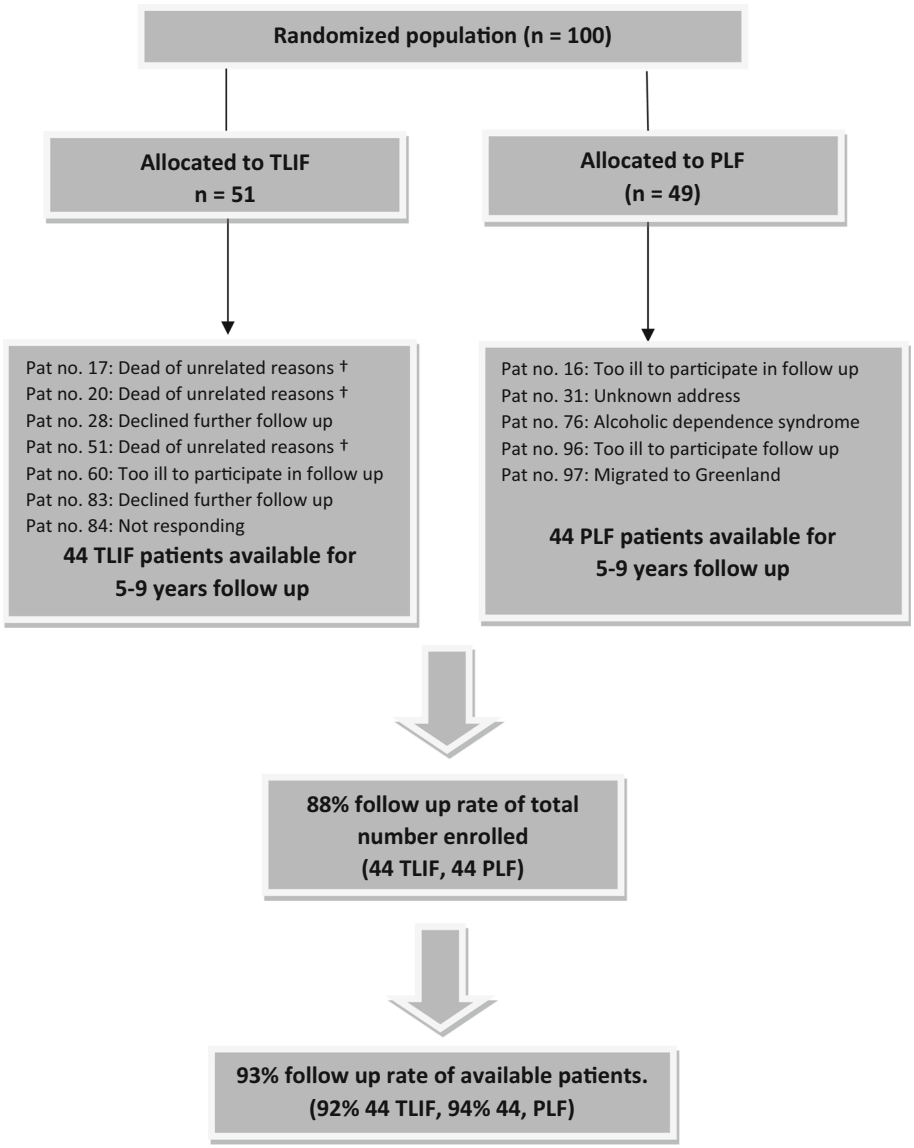


Fig. 1 Patient flow chart. Total population of 100 patients randomized in two blocks according to intervention procedure, 49 with instrumented posterolateral fusion (PLF) and 51 with transforaminal lumbar inter-body fusion (TLIF). Note: Denominator of available follow-up was with exclusions of patients with unknown address, migrated and those who had died

patients suffering from chronic low back pain. However, only one RCT with long time follow-up has been able to show benefits of this procedure [13]. In contrast, the Swedish lumbar spine study reported significantly higher complication rates in the inter-body fusion group compared to the instrumented and non-instrumented posterolateral fusion groups and no difference in functional outcome between groups [14]. Our group has earlier been able to demonstrate benefits of 360-degree fusion utilizing anterior inter-body fusion in long-term results [13]. Transforaminal lumbar inter-body fusion (TLIF) should theoretically offer the same benefits of circumferential fusion, but with less morbidity due to the lesser surgical approach. At our two 2 years follow-up of TLIF, we were not able to show any difference in functional outcome compared to a standard instrumented posterolateral fusion (PLF) [15]. However, the ALIF study [16] did not show any significant improvements at 2 years follow-up, but it changed over time in favor of ALIF at 7–9 years follow-up [13]. The aim of this study was to see whether the long-term functional outcome results of a TLIF procedure compared to a standard instrumented posterolateral fusion (PLF) could change over time in favor of anterior support.

Materials and methods

Patients

The original study was carried out between 2003 and 2008; the primary paper presents details of the methods used in this long-term follow-up [15]. Hundred patients (mean age 49.8 years) were randomized. See flow chart (Fig. 1). The regional ethical committee approved the study protocol (no:

Table 1 Demographics of patients completing follow-up

	TLIF	Control	*p*
Sex (female/male)	22/22	28/16	0.197
Age at surgery (years)	51 (30–63)	49 (25–70)	0.385
Age at follow-up (years)	59 (40–73)	57 (34–76)	0.289
Follow-up length (years)	8.6 (6–10)	8.6 (5–10)	0.796
Diagnosis			
Spondylolisthesis	10	16	0.563
DDD	17	15	
Spinal stenosis	8	6	
Failed back surgery	9	7	
Pain duration			
<1 year	6	3	0.451
1-2 years	12	16	
>2 years	26	25	
Work status			
Working	21	15	0.349
Without work	2	1	
Sick leave	12	12	
Retired	9	16	
Smoking	16 (36 %)	18 (41 %)	0.661
Previous spine surgery	9 (20 %)	13 (30 %)	0.325
Ongoing case	13 (30 %)	7 (16 %)	0.127
Operated level(s)			
1 level	32	26	0.386
2 levels	11	17	
3 levels	1	1	

Values are preoperative numbers unless stated otherwise

Values are mean (range) or number (%)

20030172). All patients suffered from severe chronic low back pain and/or leg pain, static or dynamic, resulting from the localized lumbar or lumbosacral segmental instability, spinal stenosis at levels L2-S1, or caused by isthmic spondylolisthesis (grade 1 and 2). Baseline characteristics concerning demographic and clinical data of the patients completing 5–10 years follow-up, are presented in Table 1. Before surgery, the indication for fusion was determined based on patient's history, several clinical and neurological examinations, and MRI. Exclusion criteria comprised age younger than 25 years, spondylolisthesis grade III and IV, osteoporosis diagnosed via radiography and bone mineral density testing, severe cardiac or vascular disease, brain disorders, kidney problems, former or actual malignancy, use of medicine reducing bone metabolism, dementia or abnormal psychological behavior, and language problems.

Randomization was performed using sealed envelopes, a 20-number-per block randomization. The envelopes were consecutively numbered. The type of operation remained unknown to both the patient and the surgeon until the patient's written consent was obtained.

By this procedure, initially 49 patients were allocated to posterolateral fusion with titanium TSRH (Medtronic) pedicle instrumentation (PLF group), and 51 patients to transforaminal fusion in the form of a tantalum cage (Implex/Zimmer) placed using an approach lateral to the facet joint (TLIF group), supported by a posterolateral fusion using pedicle screws (titanium TSRH, Medtronic).

All randomized patients, except for three who had died, one with unknown address, and one who migrated to Greenland were invited to an outpatients control including X-ray and MRI, see flow chart Fig. 1. So in total 95 patients were available for a 5–10 years follow-up. By that a follow-up rate 88 % of total was achieved, 93 % of available. 44 (90 % of total) 94 % of available patients in the PLF group and 44 (86 % of total) 92 % of available patients in the TLIF group.

Outcome parameters

The patients received a complementary battery of functional outcome parameters equal to those fulfilled preoperatively at 1 and 2 years follow-up including Dallas pain Questionnaire (DPQ) [17], low back pain rating scale pain index (LBPRS) [18], Oswestry disability index (ODI) [19] and the SF-36 [20]. The patients, independently of the surgeon, completed the questionnaires. Work status and employment was accessed by a questionnaire covering present employment status, questions regarding the cause of retirement concerning work status. A global outcome parameter was done using the phrase: "would you undergo the same treatment again now you know the result?".

Statistic

All analyses comparing the two intervention groups were done using the intention to treat principle. Comparison between groups were done using non-parametric tests (Mann–Whitney rank-sum test for unpaired data, Wilcoxon signed rank test for paired data, Chi square statistics or the Kruskal–Wallis test for equality at groups, with correction for ties) depending on the nature of the data. Significance level was 5 % using two-tailed testing. Intercooled Stata version 13 for Windows was the software used for statistical analysis.

Results

Functional outcome

The primary results at 2-years follow-up can be found elsewhere [15].The median time of follow-up was 8.6 years range between 5 and 10 years. The mean age of patients at follow-up was 59 years (TLIF 60/PLF 58).

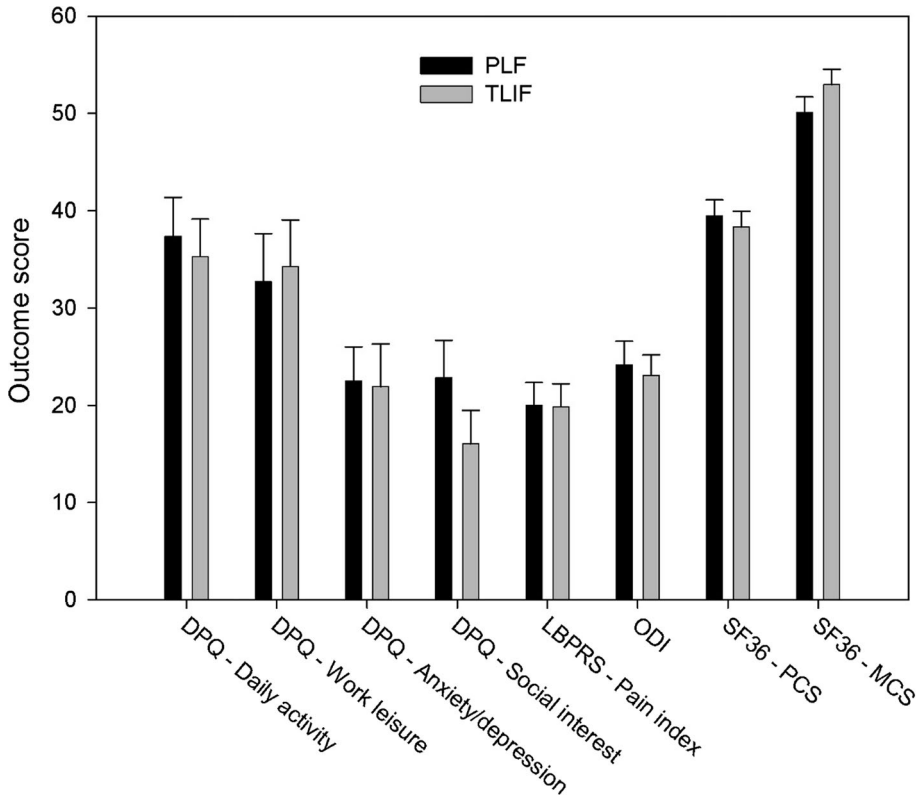


Fig. 2 Outcome score at long-term follow-up in the PLF and TLIF group. None of the differences are significant. *Bars* represent mean, *error bars* standard error of mean

The functional results in DPQ, as analyzed by means of the primary effect measure are summarized in Fig. 2. At the 5–10 years follow-up, no difference in favor of the patients in the TLIF group was found in all four categories of the DPQ. Neither were there any difference between TLIF and PLF groups in the LBPRS, ODI and SF-36 (Fig. 2). All scores remained over time statistically significant better compared to preoperative (Fig. 3).

There was no difference in mean self-reported walking distance between the two groups, although median walking distance were lower in the PLF group (TLIF: 5166/4500, PLF:4009/2000; $p = 0.131$).

Thirty-three patients (75 %) in the TLIF group and 34 patients (77 %) in the PLF group answered the question, regarding the willingness to undergo the procedure again knowing the result, positively ($p = 0.803$). During the 9 years period, a further drop in the labor force was observed, showing 28 patients still with 14 in the TLIF group and 14 in the PLF group. Whereas, 29 and 30 patients had retired from both groups, respectively ($p = 0.561$).

Reoperations

Reoperations were divided into early and late reoperations. Early reoperations was defined as immediate reoperations which occurred within the first year, such as infections, loosening of hardware, further need of decompressing. Late reoperations were defined as newly developed pathologies, such as adjacent level degeneration, which demanded further operation after 2 years. There were 11 early reoperations and 14 late reoperations. The number of early and late reoperations according to indication for reoperation is shown in Table 2. The cause of late reoperations was adjacent level segment degeneration in three and four cases, at the upper disc, next to the fusion. In two cases, the cause was disc herniation or late decompression in segments not included in the fusion. In general, reoperation was associated with significantly poorer outcomes at long-term follow-up compared to patients undergoing only their primary surgery (Fig. 4). The fusion rate was 94 % in the TLIF group, 88 % in the PLF group, $p = 0.31$ and did not change over time, from first reported, at 2 years follow-up [15].

Discussion

To our knowledge, this is the first RCT analyzing the long-term effects of TLIF in comparison to a standardized instrumented PLF. The present study is a one center study, which gives the possibility of standardizing the patient selection, the surgical techniques and the rehabilitation. With a follow-up, rate of 88 % and (93 % of available

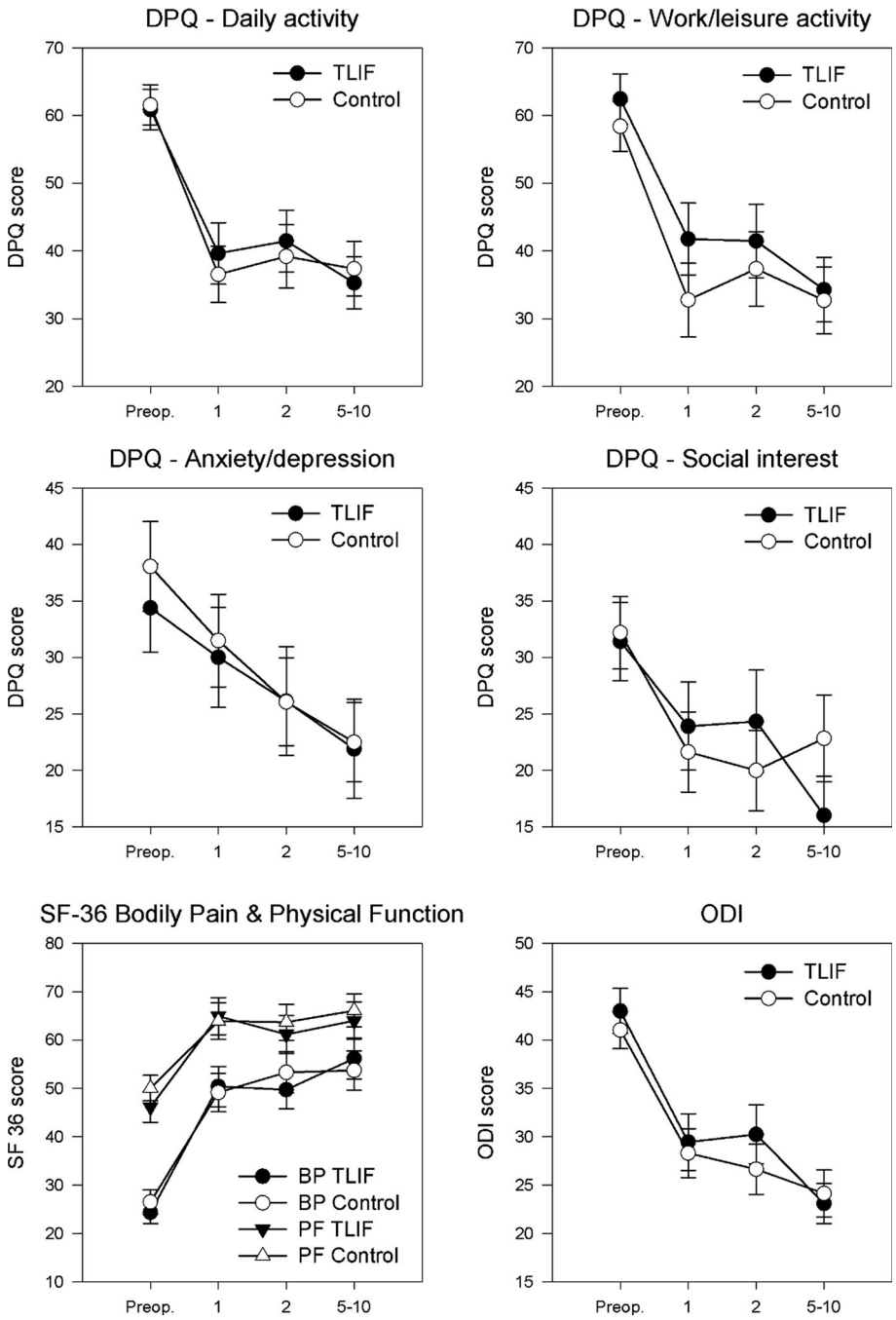


Fig. 3 Outcome scores preoperatively, at 1 and 2 years follow-up and at 5–10 year follow-up. Scores are mean, *error bars* standard error of mean. All long-term scores are significantly better than preoperative values ($p \leq 0.0305$)

inclusions) after 5–10 years (mean 8.6 years) the study can be classified as valid. Relevant issues regarding external validity are the degree of patient heterogeneity, specialization and surgical strategies and the individual surgeon's skills. Senior surgeons performed all operations. This study was done without external support from the industry, and without any disclosures connected to the randomized cohort, in that way, it can be looked upon, as totally independent of economical bias, which further increases the validity of the study.

Only a few long-term studies comparing anterior support in a prospective randomized design has been published [13, 21, 22]. However, Andersen et al. has earlier show that the positive long-term effects can be preserved for 12 years or more, which is comparable with our data regarding the PLF's, unfortunately they did not report the incidence of

Table 2 Early and late reoperations according to intervention procedure and Indication for reoperation

	PLF	TLIF	Total
Early reoperations < 1 year total	4	7	11
Hardware loosening	3	3	6
Infection	1	1	2
Early decompression	0	3	3
Disc herniation	0	0	0
Late reoperations > 2 years total	7	7	14
ASD	3	4	7
Disc herniation	1	1	2
Hardware removal	3	1	4
Late decompression	0	1	1
Total reoperations	11	14	25

The Late decompression and disc herniation was in segments not included in the fusion

reoperations in their RCT [23]. Reoperations and decline in functional outcome over time has been claimed to be due to accelerated adjacent disc degeneration (ASD) in fused segments [24, 25]. However, the decline could also be caused by aging or because of the natural history of the disease. Anterior support could result theoretical in a more lordotic fused segment, and thereby less need of compensatory lordosis, and hence, less accelerated adjacent segment degeneration and loss of gained improvement in functional outcome. In our study, the functional outcome stayed constant in both groups and significant better than preoperatively, which indicate no accelerated adjacent disc degeneration in the PLF group compared to the TLIF group. In comparison, the study from Videbæk showed significant improvement over time in the group with anterior support [13], which could advocate for the theory of improvement in lordosis due to anterior support results in better functional outcome due to less stress of the adjacent disc. However, looking into the data from Videbæk's study, the reason for significant difference between the two groups is more likely due to deterioration of the PLF group than superiority of the ALIF group, since the ODI scores in both our groups was only 24.5, and identical and comparable to their anterior support group which was 28. Unfortunately, Videbæk and all did not report numbers of late reoperations due to ASD. However, they looked into sagittal balance, lordosis and ASD, and found no sign of accelerated ASD in their PLF group, and hence, could not explain the superiority of ALIF due to less ASD due to improved lordosis and sagittal balance in their anterior support group [26, 27]. In a comparison between ALIF and TLIF regarding functional outcome, 9 studies were analyzed by Jiang et al., all studies, however, were retrospective. They did not find any difference in functional

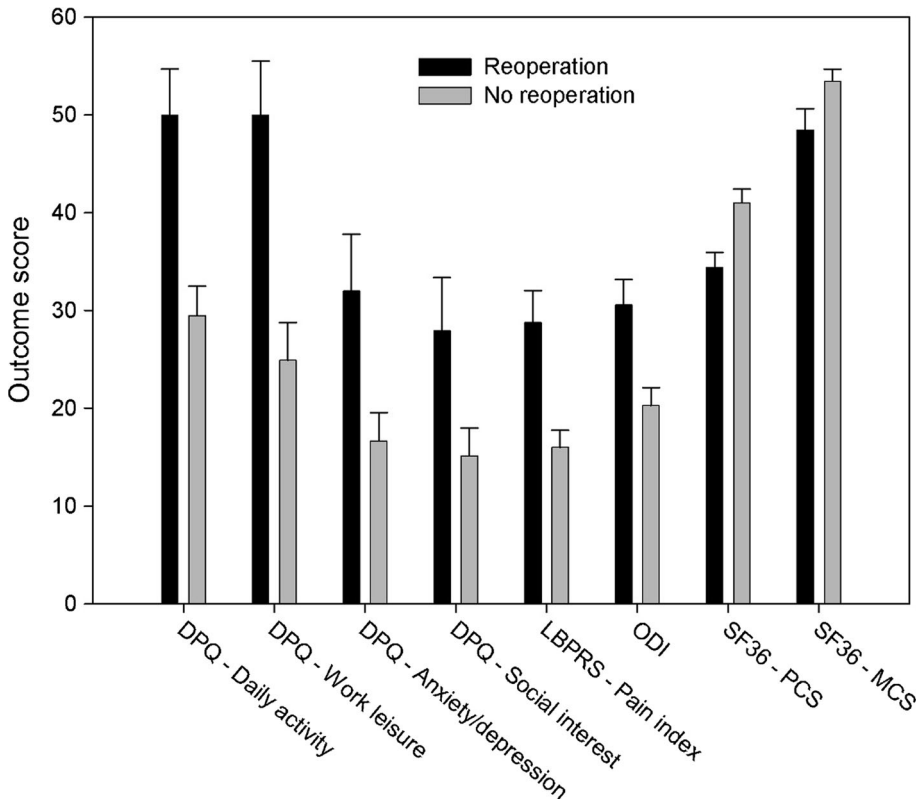


Fig. 4 Long-term follow-up outcome score comparing patients undergoing reoperation during the follow-up period to those only having the primary surgery. *Bars* represent mean, *error bars* standard error of mean. All differences significant with $p < 0.030$

outcome even though the ALIF group had better restoration of disc height, segmental lordosis, and whole lumbar lordosis [28], these findings correlate well, with our results. In addition, Audatz and colleagues were not able to find difference between three groups of patients concerning functional outcome in patients registered in a prospective manner between PLF, PLIF, TLIF [29]. Mannion et al. showed an accelerated ASD in a long-term follow-up of four controlled trials; however, it had no clinical relevance to outcome measured by the ODI. The number of dropouts was rather high, above 50 % in the surgical group and control group, furthermore neither reoperations due to ASD, nor cross over between groups were reported in the paper, which makes the study less valid. Only a single arm in one of the four RCT's the studies had anterior support [14, 22].

Another RCT's regarding TLIF focus on the difference between minimal invasive (MI-TLIF) and open TILFs. In long-term results, there is no difference between outcome in MI-TLIF and open TLIF'S [30].

TLIF has longer operation time and is more costly on bed days and production loss at 2 years follow-up [15, 31], and still does not improve functional outcome after 8.6 years, in comparison to a standard instrumented posterolateral fusion (PLF). The ODI scores are identical and comparable to the scores received in the only RCT in favor of anterior support [13]. In that aspect, it does not seem to be an alternative to instrumented PLF. However, the conflicting results still calls for further research with long-term follow-up in a prospective randomized design free of economical bias.

Compliance with ethical standards

Conflict of interest None.

References

1. Brox JI, Sorensen R, Friis A et al (2003) Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. Spine (Phila Pa 1976) 28:1913–1921
2. Brox JI, Reikeras O, Nygaard O et al (2006) Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: a prospective randomized controlled study. Pain 122:145–155
3. Moller H, Hedlund R (2000) Instrumented and noninstrumented posterolateral fusion in adult spondylolisthesis—a prospective randomized study: part 2. Spine (Phila Pa 1976) 25:1716–1721
4. Ekman P, Moller H, Hedlund R (2005) The long-term effect of posterolateral fusion in adult isthmic spondylolisthesis: a randomized controlled study. Spine J 5:36–44
5. Fritzell P, Hagg O, Wessberg P et al (2002) Chronic low back pain and fusion: a comparison of three surgical techniques: a prospective multicenter randomized study from the Swedish lumbar spine study group. Spine (Phila Pa 1976) 27:1131–1141
6. Brantigan JW, Neidre A, Toohey JS (2004) The lumbar I/F cage for posterior lumbar interbody fusion with the variable screw placement system: 10-year results of a food and drug administration clinical trial. Spine J 4:681–688
7. Harms JG, Jeszenszky D (1998) In process citation. Oper Orthop Traumatol 10:90–102
8. Gibson JN, Waddell G (2005) Surgery for degenerative lumbar spondylosis. Cochrane Database Syst Rev Cd001352
9. Thomsen K, Christensen FB, Eiskjaer SP et al (1997) 1997 Volvo Award winner in clinical studies. The effect of pedicle screw instrumentation on functional outcome and fusion rates in posterolateral lumbar spinal fusion: a prospective, randomized clinical study. Spine (Phila Pa 1976) 22:2813–2822
10. Barrick WT, Schofferman JA, Reynolds JB et al (2000) Anterior lumbar fusion improves discogenic pain at levels of prior posterolateral fusion. Spine (Phila Pa 1976) 25:853–857
11. Lee CK, Vessa P, Lee JK (1995) Chronic disabling low back pain syndrome caused by internal disc derangements. The results of disc excision and posterior lumbar interbody fusion. Spine (Phila Pa 1976) 20:356–361
12. Ekman P, Moller H, Tullberg T et al (2007) Posterior lumbar interbody fusion versus posterolateral fusion in adult isthmic spondylolisthesis. Spine (Phila Pa 1976) 32:2178–2183
13. Videbaek TS, Christensen FB, Soegaard R et al (2006) Circumferential fusion improves outcome in comparison with instrumented posterolateral fusion: long-term results of a randomized clinical trial. Spine (Phila Pa 1976) 31:2875–2880
14. Fritzell P, Hagg O, Nordwall A (2003) Complications in lumbar fusion surgery for chronic low back pain: comparison of three surgical techniques used in a prospective randomized study. A report from the Swedish Lumbar Spine Study Group. Eur Spine J 12:178–189
15. Hoy K, Bunger C, Niederman B et al (2013) Transforaminal lumbar interbody fusion (TLIF) versus posterolateral instrumented fusion (PLF) in degenerative lumbar disorders: a randomized clinical trial with 2-year follow-up. Eur Spine J 22:2022–2029
16. Christensen FB, Hansen ES, Eiskjaer SP et al (2002) Circumferential lumbar spinal fusion with Brantigan cage versus posterolateral fusion with titanium Cotrel–Dubousset instrumentation: a prospective, randomized clinical study of 146 patients. Spine (Phila Pa 1976) 27:2674–2683
17. Lawlis GF, Cuencas R, Selby D et al (1989) The development of the Dallas Pain Questionnaire. An assessment of the impact of spinal pain on behavior. Spine (Phila Pa 1976) 14:511–516
18. Manniche C, Asmussen K, Lauritsen B et al (1994) Low back pain rating scale: validation of a tool for assessment of low back pain. Pain 57:317–326
19. Fairbank JC, Pynsent PB (2000) The Oswestry disability index. Spine (Phila Pa 1976) 25:2940–2952 (**discussion 2952**)
20. Ware JE, Jr (2000) SF-36 health survey update. Spine (Phila Pa 1976) 25:3130–3139
21. Soegaard R, Bunger CE, Christiansen T et al (2007) Circumferential fusion is dominant over posterolateral fusion in a long-term perspective: cost-utility evaluation of a randomized controlled trial in severe, chronic low back pain. Spine (Phila Pa 1976) 32:2405–2414
22. Mannion AF, Leivseth G, Brox JI et al (2014) 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 (Phila Pa 1976) 39:1373–1383
23. Andersen T, Videbaek TS, Hansen ES et al (2008) The positive effect of posterolateral lumbar spinal fusion is preserved at long-

term follow-up: a RCT with 11–13 year follow-up. Eur Spine J 17:272–280
24. Levin DA, Hale JJ, Bendo JA (2007) Adjacent segment degeneration following spinal fusion for degenerative disc disease. Bull NYU Hosp Jt Dis 65:29–36
25. Park P, Garton HJ, Gala VC et al (2004) Adjacent segment disease after lumbar or lumbosacral fusion: review of the literature. Spine (Phila Pa 1976) 29:1938-1944
26. Videbaek TS, Egund N, Christensen FB et al (2010) Adjacent segment degeneration after lumbar spinal fusion: the impact of anterior column support: a randomized clinical trial with an eight- to thirteen-year magnetic resonance imaging follow-up. Spine (Phila Pa 1976) 35:1955–64
27. Videbaek TS, Bunger CE, Henriksen M et al (2011) Sagittal spinal balance after lumbar spinal fusion: the impact of anterior column support results from a randomized clinical trial with an eight- to thirteen-year radiographic follow-up. Spine (Phila Pa 1976) 36:183–191
28. Jiang SD, Chen JW, Jiang LS (2012) Which procedure is better for lumbar interbody fusion: anterior lumbar interbody fusion or transforaminal lumbar interbody fusion? Arch Orthop Trauma Surg 132:1259–1266
29. Audat Z, Moutasem O, Yousef K et al (2012) Comparison of clinical and radiological results of posterolateral fusion, posterior lumbar interbody fusion and transforaminal lumbar interbody fusion techniques in the treatment of degenerative lumbar spine. Singapore Med J 53:183–187
30. Rodriguez-Vela J, Lobo-Escolar A, Joven E et al (2013) Clinical outcomes of minimally invasive versus open approach for one-level transforaminal lumbar interbody fusion at the 3- to 4-year follow-up. Eur Spine J 22:2857–2863
31. Christensen A, Hoy K, Bunger C et al (2014) Transforaminal lumbar interbody fusion vs. posterolateral instrumented fusion: cost-utility evaluation alongside an RCT with a 2-year follow-up. Eur Spine J 23:1137–1143

Study 3:

Residual radiculopathy in TLIF surgery is not related to the side of cage insertion, but is higher in comparison to an instrumented standard posterolateral fusion measured by pain drawings. Results from and RCT

Residual radiculopathy in TLIF surgery is not related to the side of cage insertion, but is higher in comparison to an instrumented standard posterolateral fusion measured by pain drawings. Results from and RCT

Kristian Høy, Blazej Grycel, Thomas Andersen, Cody Bünger

Department of Orthopedics,

Spine Section E,

Aarhus University Hospital,

Aarhus University, Denmark.

Corresponding author:

Kristian Høy, MD

Department of Orthopedics

Spine Section

Aarhus University Hospital

Nørrebrogade 44

8000 Aarhus C,

Denmark.

Working mail: krishoey@rm.dk

Private mail: wintherhoey@dadlnet.dk

Phone +4520105058

Residual radiculopathy in TLIF surgery is not related to the side of cage insertion, but is higher in comparison to an instrumented standard posterolateral fusion measured by pain drawings. Results from a RCT

Kristian Høy, Blazej Grycel, Thomas Andersen, Cody Bünger

Department of Orthopedics,

Spine Section E,

Aarhus University Hospital,

Aarhus University, Denmark.

Corresponding author:

Kristian Høy, MD

Department of Orthopedics

Spine Section

Aarhus University Hospital

Nørrebrogade 44

8000 Aarhus C,

Denmark.

Working mail: krishoey@rm.dk

Private mail: wintherhoey@dadlnet.dk

Phone +4520105058

Word count: 3130

Abstract:

Background and purpose: TLIF has gained increasing popularity, to obtain circumferential fusion using a posterior procedure only. Cage insertion close to the exiting nerve root carries an increased risk of subsequent neurogenic pain due to manipulation of upper nerve or to the dorsal nerve root ganglion. The aim of the present study was to measure the type, appearance, localization and intensity of pain in a TLIF group compared to A PLF in a randomized study.

Patients and methods: Pain drawings and LBPRS from 100 patients (40 male, 58 female) included in a RCT comparing TLIF and PLF was analyzed. Pain drawings were completed preoperatively and at 1 and 2 year. The pain drawing consisted of a front and back outline and area under the feet.

Results :A slightly higher number of patients in the TLIF group reported any leg pain at two years follow-up: No leg pain 47% (PLF) 37% (TLIF), Unilateral leg pain 31% (PLF) 25% (TLIF), Bilateral leg pain 22% (PLF) 37% (TLIF), p=0.270.. Numbness and pins & needles on the anterior aspect of the lower leg where marked by 10% and 12% of TLIF patients compared to 6% and 4% in PLF patients p=0.498/0.197. The side of cage insertion showed no higher amount of new leg pain, compared to the contralateral side.

Interpretation: The side of cage insertion didn´t result in development of new ipsilateral leg pain; however, TLIF patients were more likely to have bilateral leg pain two years postop compared to PLF.

Introduction:

TLIF has for many years been the method of choice in order to achieve anterior support in lumbar fusions(Harms and Jeszenszky 1998, Hoy, Bunger et al. 2013, Hoy, Truong et al. 2016).The Approach changed over time from ALIFs in the 90`ties (Christensen, Hansen et al. 2002, Fritzell, Hagg et al. 2003)to TLIFs in the 00`ties. However the benefits of the anterior support in lumbar fusions has been difficult to validate (Christensen, Hansen et al. 2002, Ekman, Moller et al. 2007, Hedlund, Johansson et al. 2016, Hoy, Bunger et al. 2013, Hoy, Truong et al. 2016, Jalalpour, Neumann et al. 2015, Mannion, Brox et al. 2016). The TLIF procedure today is widely used all over the world and is now a days often done with MIS technique (Villavicencio, Burneikiene et al. 2006). The procedure carries a risk of damage to the nerve root or the dorsal root ganglion compared to a standard instrumented lumbar fusion (PLF) (Picture 1). On the other hand the use of interbody fusion might indirectly decompress the neuro foramina and by that relieving existing

radicular pain. Recently TLIF in MIS -Technique has been claimed to cause much higher rates of nerve root injuries(Epstein 2016). Due to that, we found it interesting to investigate whether this also might be a cause in open TLIF surgery versus a standardized posterolateral fusion in a randomized cohort of fused patients (PLF). One might speculate that the reason for lack of superiority in the TLIF-group (Hoy, Bunger et al. 2013, Hoy, Truong et al. 2016) could be due to cage insertion and affection, or scar tissue formation compromising the upper nerve root, resulting in residual radiculopathy or radiculitis. Pain drawings has earlier been used as a diagnostic tool (Beattie, Meyers et al. 2000, Bertilson, Brosjo et al. 2010, Mann, Brown et al. 1993, Ohnmeiss, Vanharanta et al. 1999, Ohnmeiss, Vanharanta et al. 1999, Ohnmeiss, Vanharanta et al. 1995, Prins, van der Wurff et al. 2013, Uden and Landin 1987) or in order to predict outcome in spinal lumbar surgery with conflicting results (Andersen, Christensen et al. 2010, Hagg, Fritzell et al. 2003). Initially it was developed to distinguish organic from non- organic(Uden, Astrom et al. 1988). Other studies had evaluated pain drawings to symptoms location (Beattie, Meyers et al. 2000, Bertilson, Brosjo et al. 2010, Mann, Brown et al. 1993, Ohnmeiss 2000, Ohnmeiss, Vanharanta et al. 1999). Hypothetical, pain drawings could be used as a tool, since one would expect an increase in radiating pain in the leg at the same side as the cage insertion, if the root should be irritated by the procedure or by handling or scar tissue formation (Picture 1). The aim of the present study was therefore to test the presence of pain(Margolis, Tait et al. 1986), the type of pain, the distribution of pain and the intensity of pain using pain drawings and VAS in order to show any difference in the two groups of a RCT between PLF and TLIF at 2 years follow up.

Patients and methods

The original study was carried out between 2003-2008.The primary paper presents details of the methods used in this study (Hoy, Bunger et al. 2013). 100 patients (mean age 49.8 years) were randomized. (Figure1). All patients suffered from severe chronic low back pain and/or leg pain, static or dynamic, resulting from localized lumbar or lumbosacral segmental instability, spinal stenosis at levels L2-S1 or caused by isthmic spondylolisthesis (grade 1 and 2). Baseline characteristics concerning demographic and clinical data of the patients completing pain drawings preoperative after 1 & 2 years follow up are presented in Table 1. Before surgery, the indication for fusion was determined based on anamnesis, clinical examinations CT and MRI. Exclusion criteria comprised age younger than 25 years, spondylolisthesis grade III and IV, osteoporosis diagnosed via radiography and bone mineral density testing, severe cardiac or vascular disease, brain disorders, kidney problems, former or actual malignancy, use of medicine reducing bone metabolism, dementia or abnormal psychological behavior, and language problems.

Randomization was performed using sealed envelopes, a 20-number-per block randomization. The envelopes were consecutively numbered. The type of operation remained unknown to both the patient and the surgeon until the patient's written consent was obtained.

By this procedure, initially 49 patients were allocated to posterolateral fusion with titanium TSRH (Medtronic) pedicle instrumentation (PLF group), and 51 patients to transforaminal fusion in the form of a tantalum cage (Implex/Zimmer) placed using an approach lateral to the facet joint (TLIF group), supported by a posterolateral fusion using pedicle screws (titanium TSRH, Medtronic).

Outcome parameters: Beside the pain drawings (Figure 2+3) patients received a complementary battery of functional outcome parameters equal to those fulfilled preoperatively, at 1 & 2 years follow-up including, Dallas Pain Questionnaire (DPQ)(Lawlis, Cuencas et al. 1989), Low Back Pain Rating Scale Pain Index (LBPRS)(Manniche, Asmussen et al. 1994), Oswestry Disability Index (ODI)(Fairbank and Pynsent 2000), and the SF-36(Ware 2000). The patients, independently of the surgeon, completed the questionnaires. Work status and employment was accessed by a questionnaire covering present employment status, questions regarding the cause of retirement concerning work status. A global outcome parameter was done using the phrase: "would you undergo the same treatment again now you know the result?"

The pain drawings consisted of a front and back outline of a person, as well as the area under the feet. The patients were asked to draw all their different pain using six different symbols for marking pain: 1) Dull/aching 2) burning, 3) numbness 4) Pins and needles 5) Stabbing/cutting 6) Muscular Cramps.(Figure 2). The pain drawings were divided into 11 areas using transparency overlay a then scored using a combined visual inspection and body region method as first described by Uden and Margolis et al.(Margolis, Tait et al. 1986, Uden, Astrom et al. 1988) With testing only for 1) presence and 2) type of pain. One point was allocated for the presence of one or more pain symbols within each area, resulting I a score range of 0-11. As this scoring method does not include any clinical judgement, the reproducibility is high the problem of inter-and intra-observer variability is very small. Special attention was paid to the side of insertion of the cage in the TLIF group and compared to the presence and type of newly developed pain after surgery. The data was pooled into 4 areas posterior thigh, posterior lower leg, anterior thigh and anterior lower leg and the presence of the six above mentioned types of pain in each area was registered. The second author blinded to treatment groups scored initially all pain drawings, Data was pooled and compared between preoperative and after 1 and 2 years pain drawing, as seen in table 1.The leg-pain intensity was registered using a VAS scale 0(no pain) to 10(worst) with the leg-

pain intensity experienced at the time of examination, with their average perceived leg- pain intensity in the last 14 days, and the worst leg-pain experienced in the last two weeks(Manniche, Asmussen et al. 1994).

Statistics:

All analyses comparing the two intervention groups were done using the intention to treat principle. Comparison between groups were done using non-parametric tests (Mann-Whitney rank-sum test for unpaired data, Wilcoxon signed rank test for paired data, chi-square statistics or the Kruskal-Wallis test for equality at groups, with correction for ties) depending on the nature of the data. Significance level was 5 % using two-tailed testing. Intercooled Stata version 13 for Windows was the software used for statistical analysis.

Ethics:

The local Danish regional research and ethical committee approved this study. Study protocol (No: 20030172).

Results:

Baseline between the two groups was equal Table 1. However, there was some differences in pain description in preoperative calculations, which was almost significant concerning numbness in the posterior thigh with a p value of 0.055, after 1 year this was also found between groups regarding numbness in the thigh probably due to skewed data between groups at the point of inclusion regarding numbness in the thigh. The presence and distribution of leg pain from the pooled data is shown in figure 3 and 4. The total reduction in intensity pain score from preoperative to postoperative was significant in both groups using average-leg pain VAS score within the last 14 days changed from 5.1 to 3.5 p = 0.008 (TLIF) and from 5.0 to 2.9 p=0.0003(Manniche, Asmussen et al. 1994). At 1 year, follow up five patients (10%) in the TLIF group had developed new leg pain, 2 patients on the ipsilateral side to cage insertion, 3 patients on the contralateral side. In comparison 7 patients (14%) in the control group developed new leg pain (p=0.491). TLIF patients developing pain on the side contralateral to cage insertion used following symbols: Burning (One patient), Numbness (two patients), Pins&Needles (1 patient), Cramps (1 patient). The two patients developing ipsilateral pain used: Stabbing, Dull, Pins&Needles, Burning, none of the symbols were used by both patients. Following symbols were used by the 7 patients in the control group: Numbness (5 patients), Dull (2 patients), Stabbing (2 patients), Pins&Needles (1 patient).Four of the five patients in the TLIF group, who developed leg pain at 1-year follow-up, did not mark this at their 2-year follow-up. However, another five patients marked new presence of leg pain at their 2-year follow-up in a leg marked pain free preoperatively. Two patients marked pain ipsilateral to

the side of cage insertion, two patients marked pain on the contralateral side and one patient marked the appearance of bilateral leg pain. In the control, group only one patient marked pain in a leg marked pain free pre-operatively, at 2-year follow-up. Numbness and pins & needles on the anterior aspect of the lower leg where marked by 10% and 12% of TLIF patients compared to 6% and 4% in PLF patients (p=0.498/0.157). Looking at the posterior aspect of the lower leg numbness and pins & needles where marked by 10% and 10% of TLIF patients compared to 16% and 8% in PLF patients (p=0.332/0.774). Looking at the anterior thigh the numbness pins and needle was 6% and 12 % in the TLIF group compared to 4% and 2% in the PLF group (p=o.680/0.057). The raise of new pain was further investigated according to gender, age, ongoing case and diagnosis. In none of the above mentioned categories a significant correlation could be found p=0.843, p=0.10, p=0.626

Discussion:

In this paper we looked upon pain 1 & 2 years after primary index surgery in a randomized cohort between normal posterolateral fusion (PLF) and TLIF procedure, in order to see whether there was a difference in the : 1) presence of pain, 2) type of pain, 3) distribution of pain, 4) appearance of new pain and 5) the intensity of pain. To our knowledge, this is the first study, which had used pain drawings in a randomized TLIF cohort in order to address the question whether the insertion of the cage could be a cause of development of persistent radiculitis or radiculopathy. The present study is a one- center study, which gives the possibility of standardizing the patient selection, the surgical techniques and the rehabilitation. With a follow up, rate of 94 % after 2 years, the study can be classified as valid. Relevant issues regarding external validity are the degree of patient heterogeneity, specialization and surgical strategies and the individual surgeon´s skills. Senior surgeons performed all operations. All pain drawings were done of the patient alone, beforehand and without any assistance of the doctor. The pain drawings were scored blinded to the type of intervention. This study was done without external support from the industry, and without any disclosures connected to the randomized cohort, in that way, it can be looked upon, as total independent of economical bias, which further increases the validity of the study. Pain drawings has for many years been used as a tool in order to access low back pain(Ohnmeiss 2000, Ohnmeiss, Vanharanta et al. 2000) and the repeatability concerning pain is acceptable, however the sensational drawings can be challenged. The pain drawings done in this study was a simplified version of the method used by Margolis et al.(Margolis, Tait et al. 1986) close to the simplified version used by Ohnemieis(Ohnmeiss, Vanharanta et al. 1999) this scoring method has been

shown to have very high reliability/repeatability (Lacey, Lewis et al. 2005, Margolis, Tait et al. 1986, Ohnmeiss 2000) and since it offers no clinical judgement, the scoring can be done without expert. Using this simple method, we found no difference between the two treatment groups.

Pain drawings has earlier been associated with positive radiological nerve root entrapment syndromes(Beattie, Meyers et al. 2000, Ohnmeiss, Vanharanta et al. 1999, Ohnmeiss, Vanharanta et al. 1995, Uden and Landin 1987),however Bertilson et al. found a weak agreement between MR visible nerve root entrapment and pain drawings with a specificity of 61-77%(Bertilson, Brosjo et al. 2010). Prins et al found that higher LER scores (Lower Extremity Region Scores) in pain drawings were associated with significant poorer outcome in ODI and SF(Prins, van der Wurff et al. 2013), but with insignificant prediction of appearance of poor outcome scores. These findings was in accordance to the results of Hagg et al who concluded that there was no predictive value regarding outcome measurement in the Swedish spine study. Different interpretation of the predictive values of pain drawings is however conflicting since Andersen et al. concluded that nonorganic pain drawings predicted poor functional outcome. (Andersen, Christensen et al. 2010, Hagg, Fritzell et al. 2003).

In a large multicenter study spine patient outcome research trial (SPORT), nerve root injuries occurred in 0% laminectomy of stenosis patients +/- fusion and in 2% of laminectomies due to stenosis in patients with spondylolisthesis +/- Fusion, in long term follow up this did not influence the outcome(Desai, Ball et al. 2011, Desai, Ball et al. 2012)

Our goal in the present study was, to test whether the manipulation of the nerve roots, which is mandatory in TLIF surgery, could be associated with a higher amount of radiculopathy or radiculitis, reflected in change in pain drawings and pain intensity 1 and 2 years after surgery between the randomized groups, we found no such correlation, and our result are in accordance with the results from Desai et al(Desai, Ball et al. 2011, Desai, Ball et al. 2012).

TLIF is however the method of achieving circumferential fusion with the lowest frequency of roots injuries/radiculitis 2% PLIF 7.8% and ALIFs 15.8 %(Epstein 2016, Fritzell, Hagg et al. 2003), but since the incidence of nerve root injury is higher than a normal posterolateral fusion (PLF)(Epstein 2016) the procedure should only be performed in case a posterolateral fusion is not suffice. Our results in the present study seems to support this view.

The reason of the slight rise of more pain in the TLIF group can be a reminder that lumbar fusion is not necessarily a pain removing operation, but rather a pain reducing operation and the procedure might affect the total balance of the spine (Videbaek, Bunger et al. 2011). One might speculate of the reasons. The TLIF construct is a circumferential fusion

and due to that more rigid with a higher stiffness. Theoretical this would result in earlier development of adjacent level degeneration (ADL), compared to the normal posterolateral fusion (PLF) (Cole, McCall et al. 2009) and actually this could be the primarily signs of (ADL) in our material, which calls for further research in a long term perspective. Whit regard to our results from open TLIF surgery can be looked upon as the safest method of achieving global fusion. Whether this method has a place in lumbar fusion surgery could be discussed since the results from several clinical randomized trails has not been able to show any difference in functional outcome (Christensen, Hansen et al. 2002, Fritzell, Hagg et al. 2002, Hedlund, Johansson et al. 2016, Hoy, Bunger et al. 2013, Hoy, Truong et al. 2016, Mannion, Brox et al. 2016, Mannion, Leivseth et al. 2014) . However conflicting results exist and long term results of global fusion have shown beneficial effects due to restoration of lumbar lordosis, sagittal balance and indirectly decompression of neuroforamina(Videbaek, Bunger et al. 2011, Videbaek, Christensen et al. 2006). The place and role of minimal invasive TLIF is still debatable due to the latest reports with higher and severe amounts of nerve root injuries and concomitant radiculitis (Epstein 2016, Epstein 2016). So the conclusions which can be drawn from our study is that a slighter higher percentage of the patients in the TLIF group developed leg pain 2 years after, however not significant**.** With regard to our open TLIF procedures, cage insertion could ***not*** be related to a significant increase in radiculopathy or radiculitis at ipsilateral side of cage insertion and the patients had significant reduction in pain intensity score compared to preoperative. New literature concerning TLIFs used in MIS technique has showed different results with higher nerve roots injuries and radiculitis, which indicate that, if TLIFS should be performed, it should be, performed as open procedure based of existing literature (Epstein 2016, Epstein 2016).

Contributions of authors:

KH: Study design, surgery, data collection, writing of the draft paper and revision of the paper. BG: Collection and analysis of pain drawings. TA: Data analysis, study design, revision of paper. CB: Study design, surgery.

No competing interests declared

References

Andersen, T., F. B. Christensen, K. W. Hoy, P. Helmig, B. Niedermann, E. S. Hansen and C. Bunger (2010). "The predictive value of pain drawings in lumbar spinal fusion surgery." Spine J **10**(5): 372-379.

Beattie, P. F., S. P. Meyers, P. Stratford, R. W. Millard and G. M. Hollenberg (2000). "Associations between patient report of symptoms and anatomic impairment visible on lumbar magnetic resonance imaging." Spine (Phila Pa 1976) **25**(7): 819-828.

Bertilson, B. C., E. Brosjo, H. Billing and L. E. Strender (2010). "Assessment of nerve involvement in the lumbar spine: agreement between magnetic resonance imaging, physical examination and pain drawing findings." BMC Musculoskelet Disord **11**: 202.

Christensen, F. B., E. S. Hansen, S. P. Eiskjaer, K. Hoy, P. Helmig, P. Neumann, B. Niedermann and C. E. Bunger (2002). "Circumferential lumbar spinal fusion with Brantigan cage versus posterolateral fusion with titanium Cotrel-Dubousset instrumentation: a prospective, randomized clinical study of 146 patients." Spine (Phila Pa 1976) **27**(23): 2674-2683.

Cole, C. D., T. D. McCall, M. H. Schmidt and A. T. Dailey (2009). "Comparison of low back fusion techniques: transforaminal lumbar interbody fusion (TLIF) or posterior lumbar interbody fusion (PLIF) approaches." Curr Rev Musculoskelet Med **2**(2): 118-126.

Desai, A., P. A. Ball, K. Bekelis, J. Lurie, S. K. Mirza, T. D. Tosteson and J. N. Weinstein (2011). "SPORT: does incidental durotomy affect long-term outcomes in cases of spinal stenosis?" Neurosurgery **69**(1): 38-44; discussion 44.

Desai, A., P. A. Ball, K. Bekelis, J. Lurie, S. K. Mirza, T. D. Tosteson, W. Zhao and J. N. Weinstein (2012). "Surgery for lumbar degenerative spondylolisthesis in Spine Patient Outcomes Research Trial: does incidental durotomy affect outcome?" Spine (Phila Pa 1976) **37**(5): 406-413.

Ekman, P., H. Moller, T. Tullberg, P. Neumann and R. Hedlund (2007). "Posterior lumbar interbody fusion versus posterolateral fusion in adult isthmic spondylolisthesis." Spine (Phila Pa 1976) **32**(20): 2178-2183.

Epstein, N. E. (2016). "More nerve root injuries occur with minimally invasive lumbar surgery, especially extreme lateral interbody fusion: A review." Surg Neurol Int **7**(Suppl 3): S83-95.

Epstein, N. E. (2016). "More nerve root injuries occur with minimally invasive lumbar surgery: Let's tell someone." Surg Neurol Int **7**(Suppl 3): S96-s101.

Fairbank, J. C. and P. B. Pynsent (2000). "The Oswestry Disability Index." Spine (Phila Pa 1976) **25**(22): 2940-2952; discussion 2952.

Fritzell, P., O. Hagg and A. Nordwall (2003). "Complications in lumbar fusion surgery for chronic low back pain: comparison of three surgical techniques used in a prospective randomized study. A report from the Swedish Lumbar Spine Study Group." Eur Spine J **12**(2): 178-189.

Fritzell, P., O. Hagg, P. Wessberg and A. Nordwall (2002). "Chronic low back pain and fusion: a comparison of three surgical techniques: a prospective multicenter randomized study from the Swedish lumbar spine study group." Spine (Phila Pa 1976) **27**(11): 1131-1141.

Hagg, O., P. Fritzell, R. Hedlund, H. Moller, L. Ekselius and A. Nordwall (2003). "Pain-drawing does not predict the outcome of fusion surgery for chronic low-back pain: a report from the Swedish Lumbar Spine Study." Eur Spine J **12**(1): 2-11.

Harms, J. G. and D. Jeszenszky (1998). "[In Process Citation]." Oper Orthop Traumatol **10**(2): 90-102.

Hedlund, R., C. Johansson, O. Hagg, P. Fritzell and T. Tullberg (2016). "The long-term outcome of lumbar fusion in the Swedish lumbar spine study." Spine J **16**(5): 579-587.

Hoy, K., C. Bunger, B. Niederman, P. Helmig, E. S. Hansen, H. Li and T. Andersen (2013). "Transforaminal lumbar interbody fusion (TLIF) versus posterolateral instrumented fusion (PLF) in degenerative lumbar disorders: a randomized clinical trial with 2-year follow-up." Eur Spine J **22**(9): 2022-2029.

Hoy, K., K. Truong, T. Andersen and C. Bunger (2016). "Addition of TLIF does not improve outcome over standard posterior instrumented fusion. 5-10 years long-term Follow-up: results from a RCT." Eur Spine J.

Jalalpour, K., P. Neumann, C. Johansson and R. Hedlund (2015). "A Randomized Controlled Trial Comparing Transforaminal Lumbar Interbody Fusion and Uninstrumented Posterolateral Fusion in the Degenerative Lumbar Spine." Global Spine J **5**(4): 322-328.

Lacey, R. J., M. Lewis, K. Jordan, C. Jinks and J. Sim (2005). "Interrater reliability of scoring of pain drawings in a self-report health survey." Spine (Phila Pa 1976) **30**(16): E455-458.

Lawlis, G. F., R. Cuencas, D. Selby and C. E. McCoy (1989). "The development of the Dallas Pain Questionnaire. An assessment of the impact of spinal pain on behavior." Spine (Phila Pa 1976) **14**(5): 511-516.

Mann, N. H., 3rd, M. D. Brown, D. B. Hertz, I. Enger and J. Tompkins (1993). "Initial-impression diagnosis using low-back pain patient pain drawings." Spine (Phila Pa 1976) **18**(1): 41-53.

Manniche, C., K. Asmussen, B. Lauritsen, H. Vinterberg, S. Kreiner and A. Jordan (1994). "Low Back Pain Rating scale: validation of a tool for assessment of low back pain." Pain **57**(3): 317-326.

Mannion, A. F., J. I. Brox and J. C. Fairbank (2016). "Consensus at last! Long-term results of all randomized controlled trials show that fusion is no better than non-operative care in improving pain and disability in chronic low back pain." Spine J **16**(5): 588-590.

276 Mannion, A. F., G. Leivseth, J. I. Brox, P. Fritzell, O. Hagg and J. C. Fairbank (2014). "ISSLS Prize winner: Long-term
277 follow-up suggests spinal fusion is associated with increased adjacent segment disc degeneration but without
278 influence on clinical outcome: results of a combined follow-up from 4 randomized controlled trials." Spine (Phila Pa
279 1976) **39**(17): 1373-1383.
280 Margolis, R. B., R. C. Tait and S. J. Krause (1986). "A rating system for use with patient pain drawings." Pain **24**(1): 57-
281 65.
282 Ohnmeiss, D. D. (2000). "Repeatability of pain drawings in a low back pain population." Spine (Phila Pa 1976) **25**(8):
283 980-988.
284 Ohnmeiss, D. D., H. Vanharanta and J. Ekholm (1999). "Relation between pain location and disc pathology: a study of
285 pain drawings and CT/discography." Clin J Pain **15**(3): 210-217.
286 Ohnmeiss, D. D., H. Vanharanta and J. Ekholm (1999). "Relationship of pain drawings to invasive tests assessing
287 intervertebral disc pathology." Eur Spine J **8**(2): 126-131.
288 Ohnmeiss, D. D., H. Vanharanta, A. M. Estlander and A. Jamsen (2000). "The relationship of disability (Oswestry) and
289 pain drawings to functional testing." Eur Spine J **9**(3): 208-212.
290 Ohnmeiss, D. D., H. Vanharanta and R. D. Guyer (1995). "The association between pain drawings and computed
291 tomographic/discographic pain responses." Spine (Phila Pa 1976) **20**(6): 729-733.
292 Prins, M. R., P. van der Wurff and G. J. Groen (2013). "Chronic low back pain patients with accompanying leg pain:
293 the relationship between pain extent and pain intensity, disability and health status." J Back Musculoskelet Rehabil
294 **26**(1): 55-61.
295 Uden, A., M. Astrom and H. Bergenudd (1988). "Pain drawings in chronic back pain." Spine (Phila Pa 1976) **13**(4):
296 389-392.
297 Uden, A. and L. A. Landin (1987). "Pain drawing and myelography in sciatic pain." Clin Orthop Relat Res(216): 124-
298 130.
299 Videbaek, T. S., C. E. Bunger, M. Henriksen, E. Neils and F. B. Christensen (2011). "Sagittal spinal balance after lumbar
300 spinal fusion: the impact of anterior column support results from a randomized clinical trial with an eight- to
301 thirteen-year radiographic follow-up." Spine (Phila Pa 1976) **36**(3): 183-191.
302 Videbaek, T. S., F. B. Christensen, R. Soegaard, E. S. Hansen, K. Hoy, P. Helmig, B. Niedermann, S. P. Eiskjoer and C. E.
303 Bunger (2006). "Circumferential fusion improves outcome in comparison with instrumented posterolateral fusion:
304 long-term results of a randomized clinical trial." Spine (Phila Pa 1976) **31**(25): 2875-2880.

Villavicencio, A. T., S. Burneikiene, K. R. Bulsara and J. J. Thramann (2006). "Perioperative complications in transforaminal lumbar interbody fusion versus anterior-posterior reconstruction for lumbar disc degeneration and instability." J Spinal Disord Tech **19**(2): 92-97.

Ware, J. E., Jr. (2000). "SF-36 health survey update." Spine (Phila Pa 1976) **25**(24): 3130-3139.

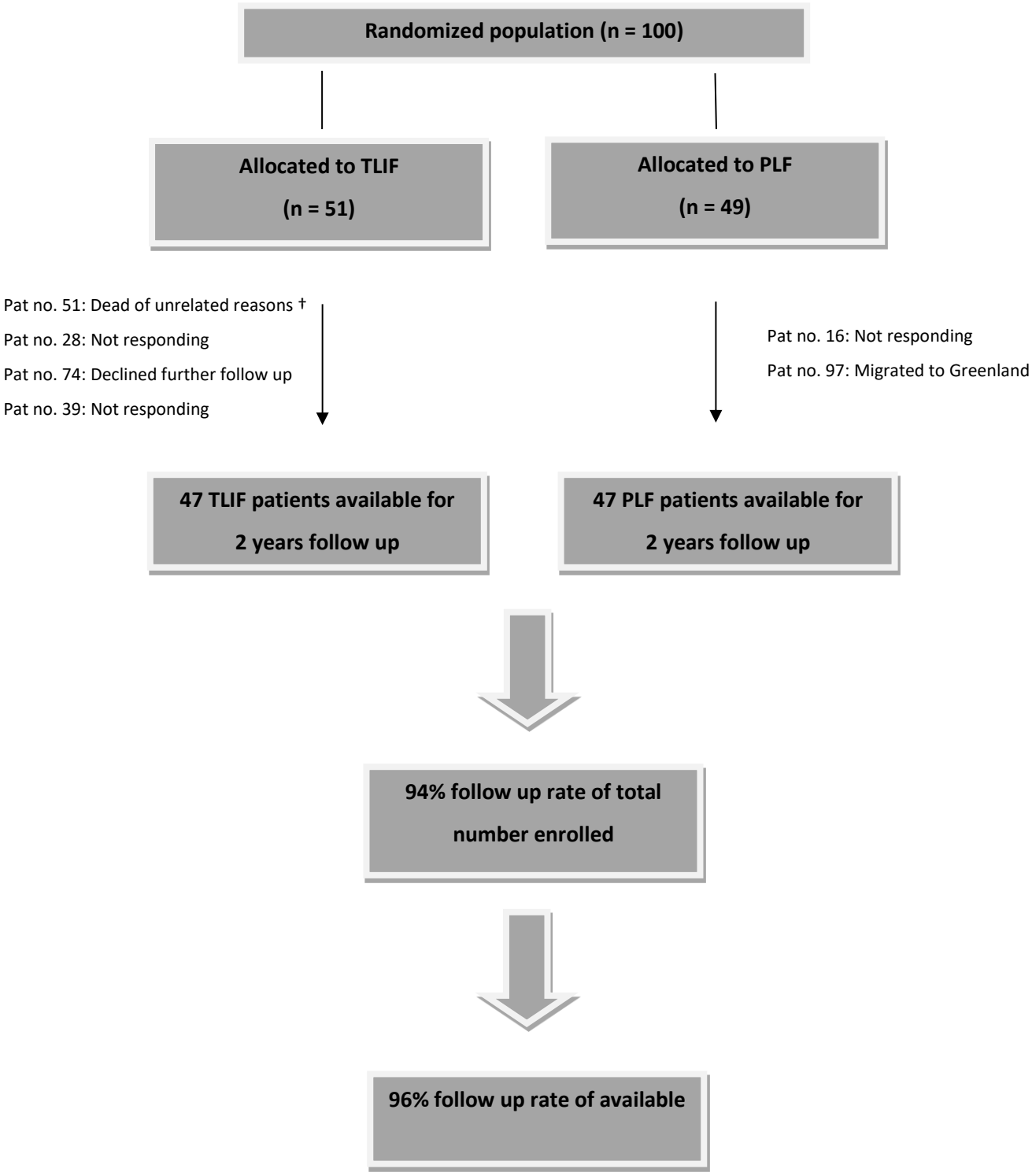


Figure 1: Patient flow chart.

Total population of 100 patients randomized in 2 blocks accordingly to intervention procedure, 49 with Instrumented Posterolateral Fusion (PLF) and 51 with Transforaminal Lumbar Interbody Fusion (TLIF)

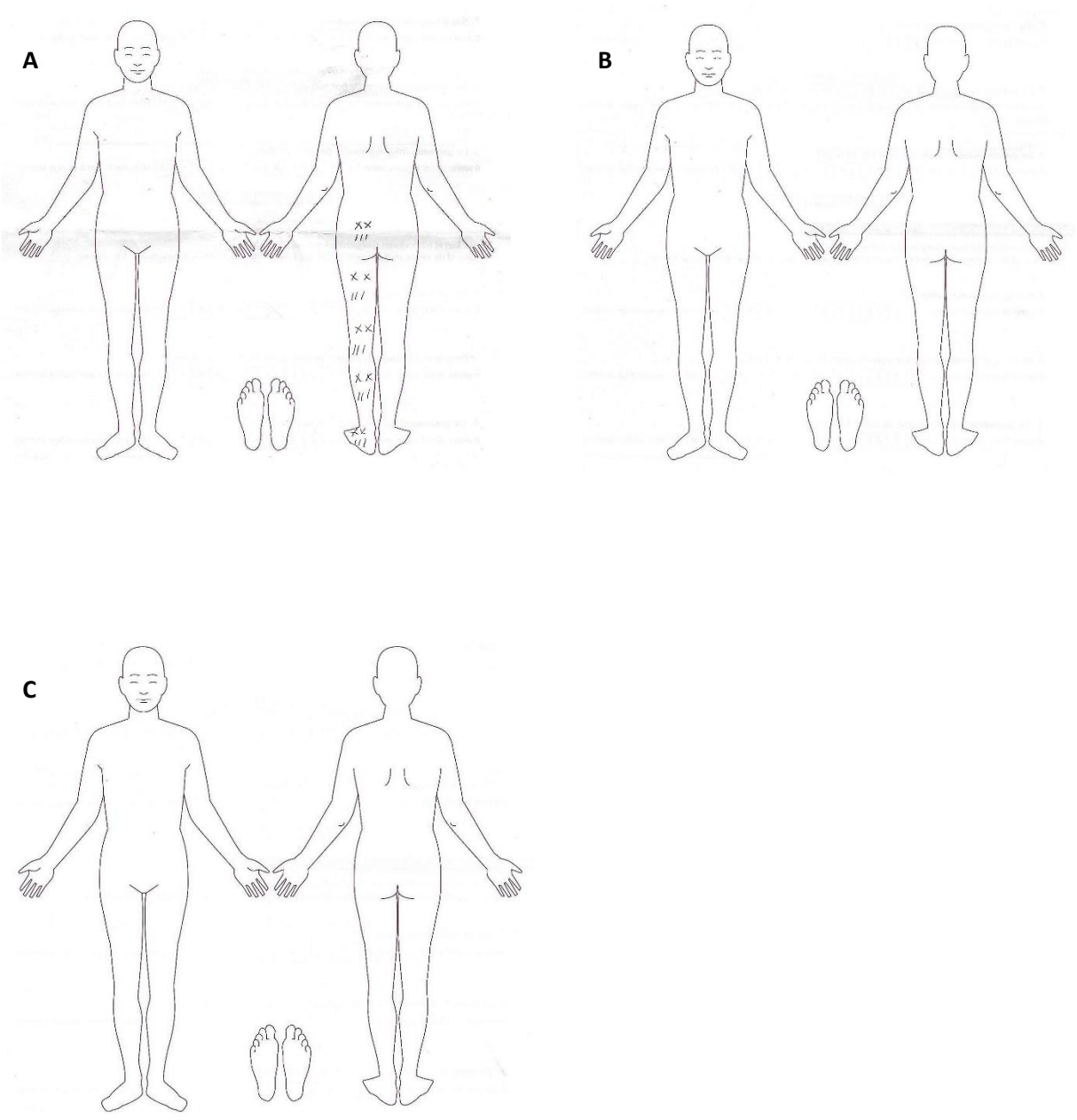


Figure 2: Representative pain drawings, example from patient number: 87 **A**: Preoperative pain drawing. Organic, S1 dermatome. **B**: Pain drawings after 1 year, pain removed, **C:** Pain drawings after 2 years Pain still removed.

Symbols used: Dull/aching *NNN*, burning *XXX*, numbness *= =*, pins and needles *:::*, stabbing/cutting */// /* and muscular cramps *SSS*.

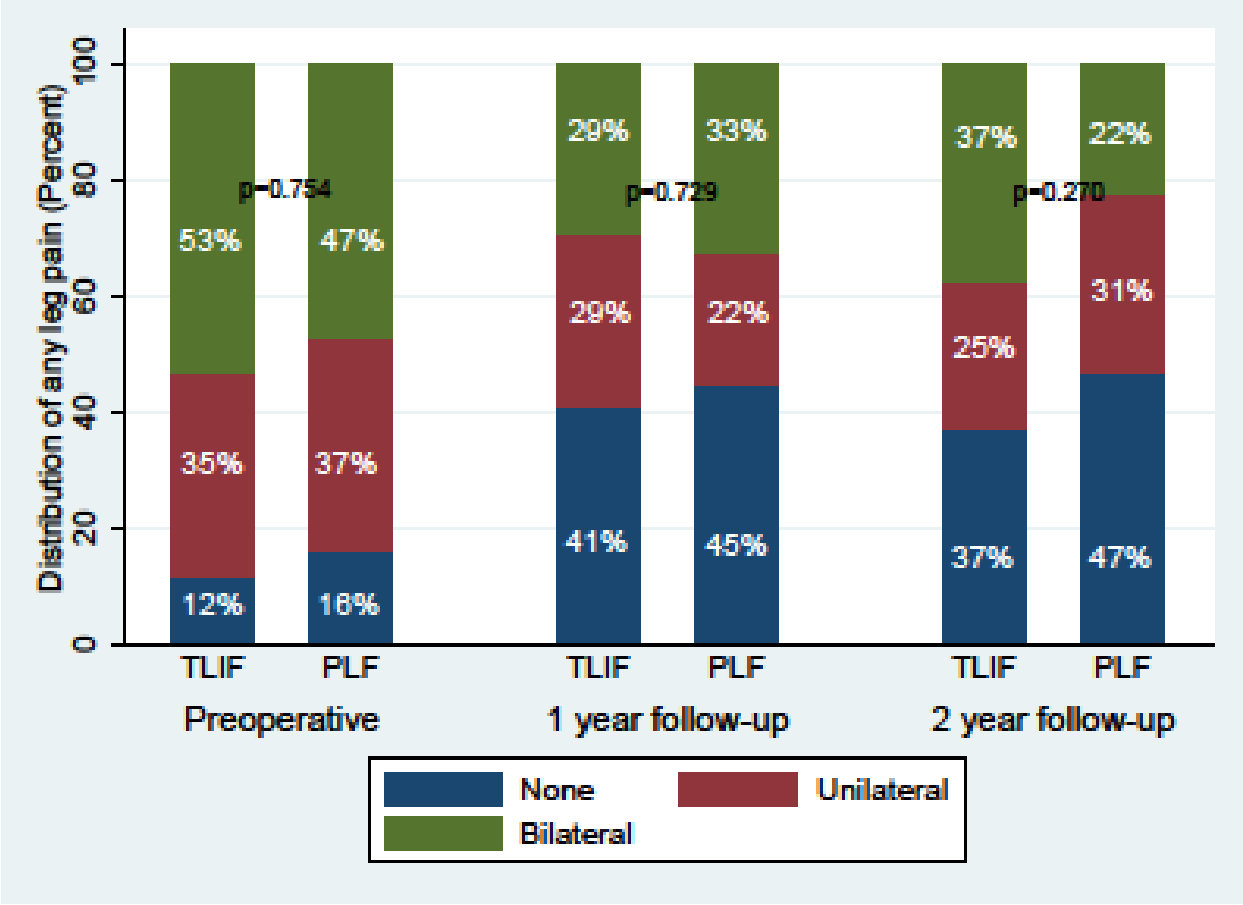


Figure 3: Presence and distribution of leg pain in the whole leg.

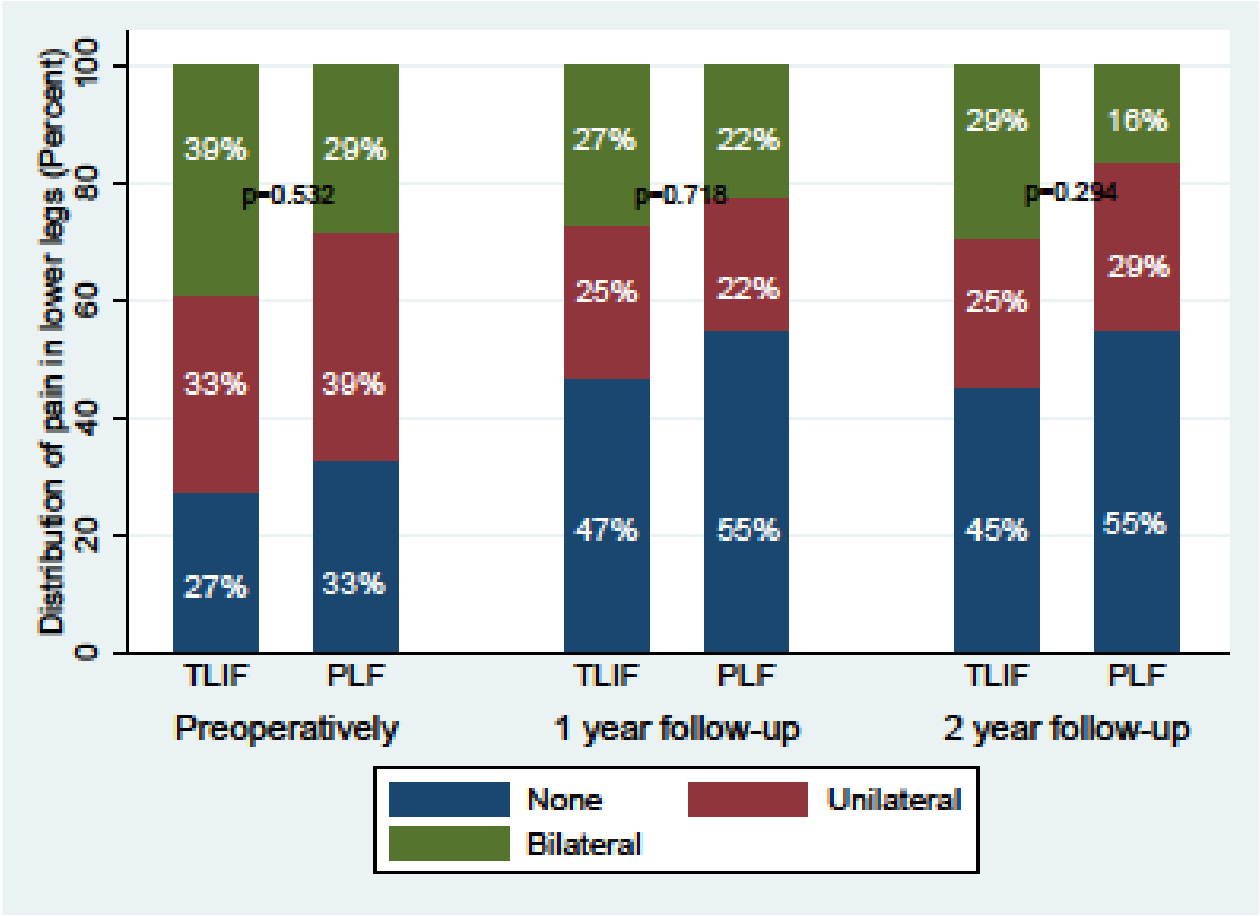
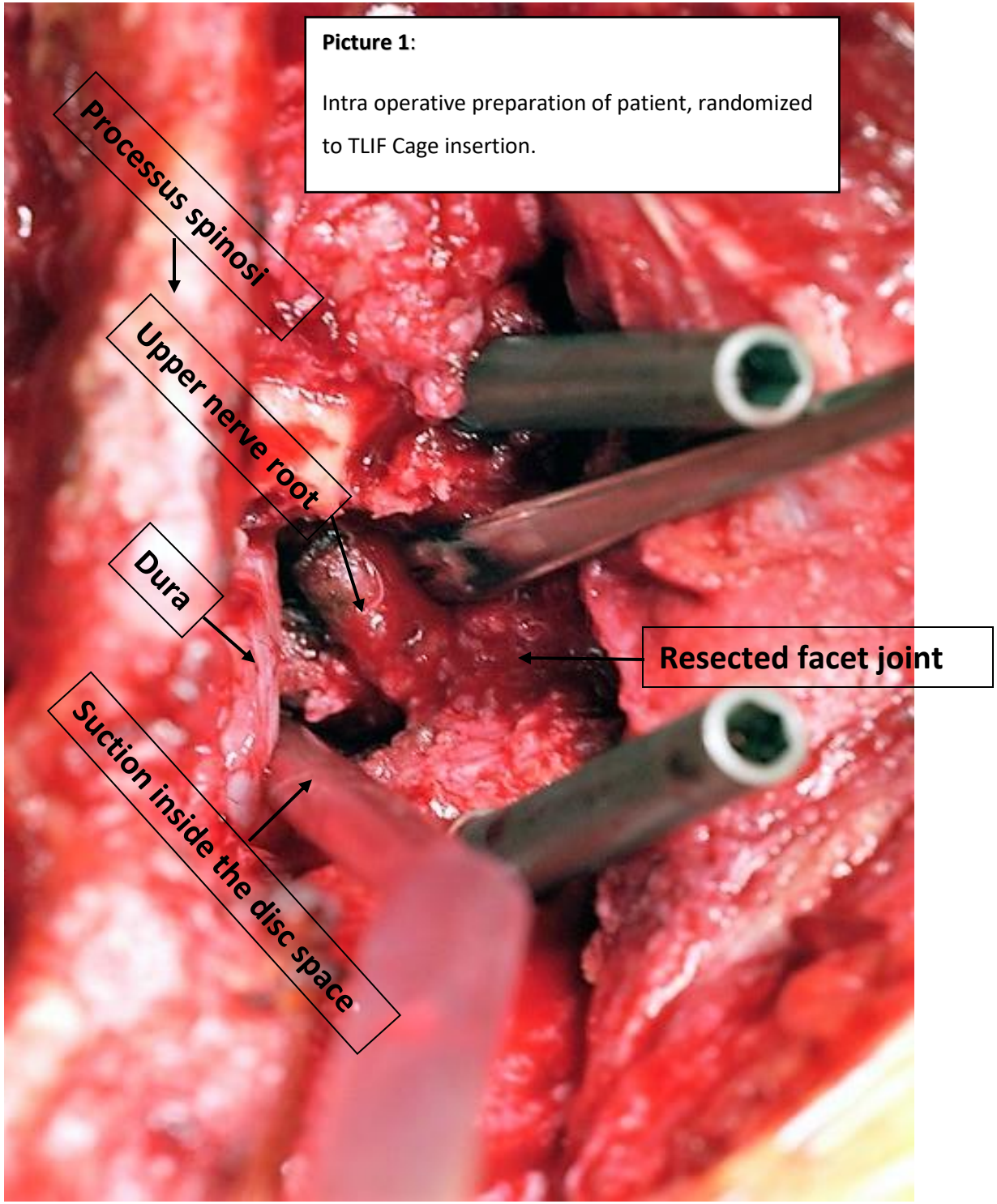


Figure 4: Presence and distribution of leg pain below the knee



Picture 1:

Intra operative preparation of patient, randomized to TLIF Cage insertion.

Tabel 1: Presence and type of pain in the 4 pooled areas.

	Preoperatively			1 year follow-up			2 year follow-up		
	TLIF	**PLF**	p-value	**TLIF**	**PLF**	p-value	**TLIF**	**PLF**	p-value
Posterior thigh									
Dull/aching	11 (22%)	11 (22%)	0.915	2 (4%)	9 (18%)	0.021	6 (12%)	2 (4%)	0.157
Stabbing/cutting	14 (27%)	11 (22%)	0.564	4 (8%)	4 (8%)	0.953	4 (8%)	4 (8%)	0.953
Burning	10 (20%)	6 (12%)	0.315	1 (2%)	3 (6%)	0.288	1 (2%)	0 (0%)	0.325
Numbness	3 (6%)	9 (18%)	0.055	0 (0%)	7 (14%)	0.005	1 (2%)	5 (10%)	0.083
Pins & needles	4 (8%)	8 (16%)	0.192	2 (4%)	4 (8%)	0.372	6 (12%)	4 (8%)	0.548
Muscular cramps	4 (8%)	5 (10%)	0.680	2 (4%)	2 (4%)	0.967	3 (6%)	3 (6%)	0.960
Posterior lower leg									
Dull/aching	14 (27%)	7 (14%)	0.106	6 (12%)	6 (12%)	0.941	5 (10%)	3 (6%)	0.498
Stabbing/cutting	10 (20%)	7 (14%)	0.479	4 (8%)	5 (10%)	0.680	5 (10%)	5 (10%)	0.947
Burning	9 (18%)	4 (8%)	0.159	3 (6%)	2 (4%)	0.680	2 (4%)	2 (4%)	0.967
Numbness	14 (27%)	12 (24%)	0.736	6 (12%)	10 (20%)	0.239	5 (10%)	8 (16%)	0.332
Pins & needles	13 (25%)	11 (22%)	0.722	5 (10%)	5 (10%)	0.947	5 (10%)	4 (8%)	0.774
Muscular cramps	10 (20%)	10 (20%)	0.920	10 (20%)	4 (8%)	0.099	9 (18%)	5 (10%)	0.284
Anterior thigh									
Dull/aching	11 (22%)	8 (16%)	0.504	5 (10%)	6 (12%)	0.697	10 (20%)	5 (10%)	0.188
Stabbing/cutting	8 (16%)	6 (12%)	0.620	1 (2%)	5 (10%)	0.083	4 (8%)	3 (6%)	0.736
Burning	6 (12%)	5 (10%)	0.803	7 (14%)	3 (6%)	0.205	5 (10%)	1 (2%)	0.102
Numbness	3 (6%)	2 (4%)	0.680	2 (4%)	6 (12%)	0.125	3 (6%)	2 (4%)	0.680
Pins & needles	4 (8%)	6 (12%)	0.463	4 (8%)	2 (4%)	0.428	6 (12%)	1 (2%)	0.057
Muscular cramps	2 (4%)	1 (2%)	0.582	0	0	NA	4 (8%)	1 (2%)	0.183
Anterior lower leg									
Dull/aching	6 (12%)	4 (8%)	0.548	5 (10%)	2 (4%)	0.262	4 (8%)	7 (14%)	0.303
Stabbing/cutting	4 (8%)	6 (12%)	0.463	4 (8%)	6 (12%)	0.463	5 (10%)	1 (2%)	0.102
Burning	5 (10%)	3 (6%)	0.498	4 (8%)	3 (6%)	0.736	2 (4%)	1 (2%)	0.582
Numbness	5 (10%)	7 (14%)	0.491	5 (10%)	7 (14%)	0.491	5 (10%)	3 (6%)	0.498
Pins & needles	4 (8%)	8 (16%)	0.192	7 (14%)	3 (6%)	0.205	6 (12%)	2 (4%)	0.157
Muscular cramps	6 (12%)	4 (8%)	0.548	4 (8%)	0 (0%)	0.045	9 (18%)	4 (8%)	0.159

No patients used more than 2 symbols at one time at any stage

Table 2			
	Pre-operatively	1 year follow-up	2 year follow-up
No leg pain	6 (12%)	21 (41%)	19 (37%)
Ipsilateral leg pain	13 (25%)	8 (16%)	9 (18%)
Contralateral leg pain	5 (10%)	7 (14%)	4 (8%)
Bilateral leg pain	27 (53%)	15 (29%)	19 (37%)

Table 2: Distribution of leg pain according to side of TLIF cage insertion. Difference between time point are significant comparing preoperatively with 1 year follow-up (p=0.004) and 2 year follow-up (p=0.029), but with no difference between 1 and 2 year follow-up (p=0.694)

Study 4:

Transforaminal lumbar interbody fusion vs. posterolateral instrumented fusion: cost-utility evaluation along side an RCT with a 2-year follow-up

Eur Spine J (2014) 23:1137–1143
DOI 10.1007/s00586-014-3238-6

ORIGINAL ARTICLE

Transforaminal lumbar interbody fusion vs. posterolateral instrumented fusion: cost-utility evaluation along side an RCT with a 2-year follow-up

A. Christensen · K. Høy · C. Bünger · P. Helmig · E. S. Hansen · T. Andersen · R. Søgaard

Received: 30 August 2013 / Revised: 5 February 2014 / Accepted: 6 February 2014 / Published online: 21 February 2014

Abstract
Purpose Long-lasting low back pain is an increasing problem, and for some patients surgery is the final option for improvement. Several techniques for spinal fusion are available and the optimal technique remains uncertain. The objective of this study was to assess the cost-effectiveness and cost-utility of transforaminal lumbar interbody fusion (TLIF) compared to posterolateral instrumented fusion (PLF) from the societal perspective.
Methods 100 Patients were randomized to TLIF or PLF (51/49) and followed for 2 years. Cost data were acquired from national registers, and outcomes were measured using the Oswestry Disability Index and SF-6D questionnaires. Conventional cost-effectiveness methodology was employed to estimate net benefit and to illustrate cost-effectiveness acceptability curves. The statistical analysis was based on means and bootstrapped confidence intervals.
Results Results showed no statistically significant difference in either cost or effects although a tendency for the TLIF regimen being more costly on bed days (€2,554) and production loss (€1,915) was observed. The probability that TLIF would be cost-effective did not exceed 30 % for any threshold of willingness to pay per quality-adjusted life year. Sensitivity analysis was conducted and supported the statistical model for handling of missing data.
Conclusion TLIF does not seem to be a relevant alternative to PLF from a socioeconomic, societal point of view.

Keywords RCT · Economics · Cost-effectiveness · Cost-utility · Transforaminal lumbar interbody fusion · Posterolateral fusion

Introduction

Low back pain has become a major health problem during the last couple of decades, and it has been estimated that up to 23 % of low back pain cases become chronic [1]. Often, therapeutic treatment can relieve pain and restore functional ability but in some cases surgery might be the ultimate solution.

Several surgical techniques for stabilization of the spine have been developed and tested in randomized designs. One of these is transforaminal lumbar interbody fusion (TLIF), which has been compared to various forms of posterolateral fusion (PLF) without convincing clinical benefits [2–5]. However, it was observed in three of these studies that the TLIF strategy led to fewer complications in the postoperative phase [3–5], which, in a socioeconomic perspective, could translate into less health care use and faster return to work. It, thus, remains uncertain whether TLIF is a cost-effective procedure due to a benefit in socioeconomic parameters.

A. Christensen (✉)
Centre for Applied Health Services Research, University of Southern Denmark, J. B. Winsløw Vej 9.B, 5000 Odense C, Denmark
e-mail: ann.christensen3@gmail.com

K. Høy · C. Bünger · P. Helmig · E. S. Hansen · T. Andersen
Spine Section, Department of Orthopedics E, Aarhus University Hospital, Aarhus C, Denmark

R. Søgaard
Public Health and Quality Improvement, Central Denmark Region, Aarhus N, Denmark

R. Søgaard
Institute of Public Health, Aarhus University, Aarhus, Denmark

There seem to be no economic evaluations of TLIF in the literature although several studies have been conducted for other techniques for spinal fusion [6, 7]. Some of the surgical techniques are often referred to as expensive due to high index treatment costs, but it should be noted that in the societal perspective enabling just one extra patient to return to work can quickly outweigh the higher hospital costs of several patients.

The present study was conducted alongside a randomized controlled trial assessing the clinical efficacy of TLIF versus instrumented PLF [2]. Functional ability was found to improve significantly in both groups over the 2-year follow-up, but no significant difference between the groups could be established. Apart from being the first economic evaluation addressing the TLIF technique, the motivation for this study is to inform the remaining uncertainty about the attractiveness of the TLIF technique; whether it could be clinically inefficient but still cost-effective due to less service utilization and/or less production loss to society. Thus, the objective of this study was to assess the cost-effectiveness and the cost-utility of TLIF versus instrumented PLF in a societal perspective.

Materials and methods

Study design and population

This study was conducted alongside an RCT with a 2-year follow-up period. 100 patients were included in the study between November 2003 and November 2008, and randomization was done using sealed envelopes with a 20-number per block randomization. 51 patients were randomized to TLIF and 49 were randomized to PLF.

The study population included patients with a history of long-lasting low back pain at a specialized university hospital clinic. Patients with the following characteristics were ineligible: age <25 years, spondylolisthesis grade III and IV, osteoporosis, severe cardiac or vascular disease, brain disorders, kidney problems, former or actual malignancy, use of medicine reducing bone metabolism, dementia or abnormal psychological behavior, and language problems. Further description of the study population has been previously reported [2].

The study was approved by the Danish Data Protection Agency (2013-41-1437), and the Regional Ethical Committee (20030172).

Economic evaluation

A combined cost-effectiveness and cost-utility evaluation was conducted according to standard methodology [8]. Costs and effects were discounted at an annual rate of 3 % before the incremental cost-effectiveness ratio (ICER) was calculated using the formula: $\text{ICER} = (C_{\text{TLIF}} - C_{\text{PLF}})/(E_{\text{TLIF}} - E_{\text{PLF}})$. This ratio cannot be interpreted if any of the parameters are negative and costs and effects were therefore converted into net benefit and presented in so-called cost-effectiveness acceptability curves [9]. These curves illustrate the probability that an intervention (TLIF) is cost-effective compared to the control (PLF) for a continuum of hypothetical threshold values of willingness to pay per unit of effect.

Costing

The societal perspective was made up of health care use in primary care, in hospitals and productivity losses due to inability to participate in the labor force. All costs were converted into euros (€) and inflated to the common price year 2012 using the consumer price index where relevant.

Data on primary health care use (number of visits and the related activity-based tariffs) were extracted from The National Health Insurance Service Register [10], while data on use of secondary health care services [number of services and national average Diagnosis-Related Grouping (DRG) tariffs] were extracted from the National Patient Registry [11] and the DRG registry [12]. The DRG tariff is revised every year and is based on average use of resources in each DRG group. The DRG tariffs include examination, bed days, surgery and postoperative control. A great advantage of using administrative register data is the 100 % response on both the service use and how to value it. However, no tariffs were available for the years 2003 and 2004, which therefore had to be valued manually. This was the case for the index surgery of 22 patients from the TLIF group and 19 from the PLF group. The manual valuation was conducted using 2012 tariffs.

The DREAM database, containing information on all social benefits, was searched for events of inability to work. Productivity losses were calculated using weeks of inability to work for those who were part of the labor force at baseline and not reaching the age for pension during follow-up. Age- and gender-matched average gross salaries [13] were used to value production losses.

Outcome parameters

The primary outcome parameter was health-related quality of life as defined by the SF-6D [14] while the Oswestry Disability Index (ODI) [15] was used as a secondary outcome parameter. The SF-6D observations were preference-weighted using British weights [14] to estimate quality-adjusted life years (QALYs).

Imputation

The data suffered missing values on the patient-reported SF-6D and the ODI, while a 100 % response was acquired on the cost parameters. To avoid loss of information, the missing values on the outcome parameters were imputed by a mean within the randomization group. Furthermore, sensitivity analysis was conducted for the alternative analytical choice of carrying the 1-year follow-up observation forward to impute missing values on 2-year follow-up.

Statistical methods

Baseline characteristics were summarized using conventional summary statistics. The *p* values were derived using Wilcoxon rank sum test.

Comparative analysis was based on means with bootstrapped standard errors (10,000 replications were drawn) [16]. This method was applied, as resource use, costs and outcomes were highly skewed and therefore not applicable for parametric statistics.

The analytical strategy was implemented for two scenarios: based on cases with complete response on the effect parameters and based on an imputed data set where missing values on the effect parameters were imputed. Although both scenarios are shown, the latter was considered as the main analysis.

Results

Population characteristics

The study population included 59 % females and was characterized by a mean age of 50 years (see Table 1). No statistical differences were observed on baseline characteristics between randomization groups.

The response rate at the 2-year follow-up was 90 and 88 % in the TLIF and the PLF groups, respectively. A non-response analysis was made for age, gender, work status, primary health care cost, secondary health care cost, production loss and total cost for missing values in effect parameters. No significant difference was found except that SF-6D and ODI non-responders appeared to have generated lower production loss than responders. This difference carried over to total costs and indicated that non-responders were in fact better off than responders. The extent to which non-responders differed from responders was the same in the two randomization groups.

Resource use

The PLF group seemed to have had more visits to the general practitioner and therapists compared to the TLIF group (see Table 2). Opposite, the TLIF group had approximately two outpatient visits and about 4 days of sick leave more than the PLF group, although none of these results were statistically significant.

Table 1 Demographics and status at baseline

	TLIF ($n = 51$)	PLF ($n = 49$)	p value
Age at surgery (mean and range)	51 (30–63)	49 (25–70)	0.27
Male gender	24 (47)	17 (35)	0.21
Work status			0.15
Working	18 (35)	11 (23)	
Work shorter hours	4 (8)	4 (8)	
Rehabilitation	0	1 (2)	
Incapacity benefits[a]	2 (4)	0	
Sick leave	14 (27)	15 (31)	
Without work	2 (4)	2 (4)	
Early retirement	0	3 (6)	
Retired	11 (22)	13 (27)	
ODI (mean and range)	44 (12–86)	40 (18–66)	0.12
SF-6D baseline (mean and range)	0.55 (0.30–0.92)	0.56 (0.40–0.81)	0.63

Values are in numbers (percentage) unless otherwise stated

[a] Incapacity benefits due to back problems

The biggest cost difference between the groups related to bed days and production losses. There was no difference between the two groups on the number of bed days, but when these were valued according to the activity taking place during admissions, the TLIF group was on average €2554 more expensive than the PLF group, indicating that they had more (costly) procedures during their admissions on top of the extra cost associated with the TLIF procedure per se. As concerns production losses, the inability to work of the TLIF group translated into an excess production loss of €1915 as compared with the PLF group Table 3.

Health outcomes

The two effect parameters demonstrated that both groups improved significantly from baseline to 12 months after surgery, but none or only little improvement was demonstrated from 12 to 24 months after surgery (see Table 4). No overall significant difference was observed between the randomization groups, neither on the QALY measure nor on the ODI.

Cost-effectiveness and cost-utility

Figure 1 shows the probability that TLIF is cost-effective compared to PLF at a given threshold for the maximum willingness to pay per QALY. At a willingness to pay as high as €100,000 there is only 22 % probability that TLIF

Table 2 Use of resources during a 2-year follow-up period

	TLIF ($n = 51$)	PLF ($n = 49$)	Diff.	95 % CI for difference (bootstrapped)	p value (bootstrapped)
Primary health care					
General practitioner	31.96	36.22	−4.26	(−18.54; 10.01)	0.55
Medical specialist	1.69	1.61	0.07	(−1.21; 1.36)	0.91
Therapist	1.35	2.67	−1.32	(−2.75; 0.11)	0.07
Dentistry	2.82	2.98	−0.16	(−0.93; 0.62)	0.69
Other	0.06	0.02	0.04	(−0.04; 0.11)	0.33
Secondary health care					
Bed days	14.29	14.22	0.07	(−4.45; 4.59)	0.91
Outpatient visits	8.73	6.86	1.87	(−1.27; 5.00)	0.24
Emergency room	0.10	0.02	0.08	(−0.01; 0.17)	0.10
Sick leave					
Weeks	34.20	33.47	0.73	(−13.66; 15.11)	0.92

Values are shown as mean units per patient

Table 3 Resource use from baseline through 2-year follow-up

	TLIF ($n = 51$)	PLF ($n = 49$)	Diff.	95 % CI for difference (bootstrapped)	p value (bootstrapped)
Primary health care					
General practitioner	469	586	−117	(−425.23; 192.33)	0.45
Medical specialist	204	146	58	(−112.29; 228.22)	0.50
Therapist	126	97	28	(−90.82; 146.95)	0.64
Dentistry	116	122	−6	(−42.73; 30.46)	0.74
Other	2	1	−1	(−3.30; 3.71)	0.91
Primary health care costs	916	953	−37	(−438.51; 364.81)	0.86
Secondary health care					
Bed days	22,267	19,713	2554	(−698.90; 5807.18)	0.12
Out-patient visits	1,647	1,428	220	(−426.34; 865.36)	0.51
Emergency room visits	13	3	11	(−1.80; 22.94)	0.10
Secondary health care costs	23,929	21,144	−2784	(−771.15; 6339.60)	0.13
Sick leave					
Production loss costs	35,019	33,103	1915	(−13,253.10; 17,085.55)	0.81
Total costs	59,863	55,200	4663	(−10,919.07; 20,245.26)	0.56

Values are mean cost per patient in Euros

is cost-effective over PLF. This was also the case for the sensitivity analyses based on complete-response cases only and based on the alternative imputation strategy of last observation carried forward, respectively.

Figure 2 shows the corresponding probability that TLIF is cost-effective compared to PLF per 1-point gain on the ODI scale and it can be seen that quite high thresholds are needed to achieve a clinically relevant ODI gain.

Discussion

It has already been established that there is no major difference on the functional outcome of TLIF and PLF, respectively. Several studies have however indicated that the TLIF procedure could lead to less complications [3–5], which could make this strategy cost-effective, even though it is not clinically superior, due to fewer health care costs and/or less productivity loss. None of these consequences could be demonstrated in the present setting, and the overall conclusion therefore seems to be quite clear that TLIF is not a cost-effective alternative to instrumented PLF.

Comparison with other economic evaluations

To the best of our knowledge, this is the first study on the specific technique of TLIF whereas other studies have

Table 4 Health outcome during 2-year follow-up

	TLIF	PLF	Diff.	95 % CI for difference (bootstrapped)	*p* value (bootstrapped)
ODI					
Baseline					
Complete case (40/32)	44	40	4	(−2.84; 10.97)	0.25
Imputed case (51/49)	44	40	4	(−0.99; 9.12)	0.12
12 months					
Complete case (32/36)	29	27	2	(−7.27; 11.36)	0.67
Imputed case (51/49)	29	27	2	(−4.11; 8.21)	0.51
24 months					
Complete case (34/32)	29	25	3	(−5.89; 12.58)	0.48
Imputed case (51/49)	29	25	3	(−2.82; 9.51)	0.29
ODI difference					
Complete case (28/25)	16	13	4	(−5.47; 12.48)	0.44
Imputed case (51/49)	17	16	1	(−5.04; 6.67)	0.78
SF-6D					
Baseline					
Complete case (45/42)	0.55	0.56	−0.01	(−0.06; 0.03)	0.57
Imputed case (51/49)	0.55	0.56	−0.01	(−0.05; 0.03)	0.63
12 months					
Complete case (43/44)	0.66	0.66	0.00	(−0.06; 0.07)	0.89
Imputed case (51/49)	0.66	0.66	0.00	(−0.05; 0.06)	0.89
24 months					
Complete case (45/39)	0.66	0.67	−0.02	(−0.08; 0.05)	0.57
Imputed case (51/49)	0.66	0.67	−0.01	(−0.07; 0.04)	0.63
QALYs					
Complete case (39/34)	1.24	1.27	−0.03	(−0.15; 0.09)	0.61
Imputed case (51/49)	1.23	1.26	−0.03	(−0.11; 0.06)	0.60

Values are shown in mean. Imputed case is used as the main analysis

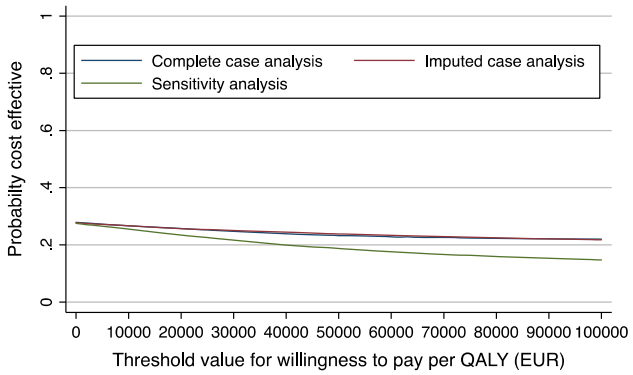


Fig. 1 Cost-effectiveness acceptability curve showing the probability that TLIF is cost-effective over PLF, based on the outcome in terms of quality-adjusted life years (QALY). The curve shows that the probability, that TLIF is cost-effective over PLF, decreases as threshold value for willingness to pay increases

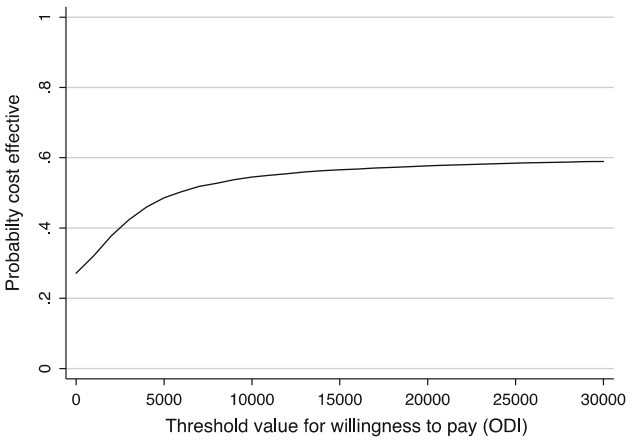


Fig. 2 Cost-effectiveness acceptability curve showing the probability that TLIF is cost-effective over PLF based on the outcome measured by the Oswestry Disability Index. The curve shows that the probability, that TLIF is cost-effective over PLF, increases along with the threshold value for willingness to pay per ODI point

assessed the cost-effectiveness of different techniques for lumbar spinal fusion [17–22].

Four of the six reported studies found no significant difference in functional outcome or QALYs between intervention and control group [17–19, 22]. Freeman et al. found that use of the standard femoral ring allografts when performing circumferential lumbar fusion was both cheaper, generated more QALYs and had a higher return to work rate compared to the intervention treatment using titanium cages [21]. Søgaard et al. also found a significant difference in functional outcome (using EQ-5D) and costs when concluding that circumferential fusion was superior to posterolateral fusion [20].

TLIF was insignificantly more expensive on bed days, indicating that these patients received more services during their admission compared to the PLF patients.

The grand total difference between the groups was a non-significant difference of €4663 with the TLIF group being the most costly while nearly producing the same amount of QALYs and indicating that TLIF produced more leg pain. The present results are, therefore, not in consensus with the studies finding TLIF to be superior to other lumbar fusion techniques [3–5].

Strengths and weaknesses of the present study

The randomized study design is the strongest for assessing differences between groups. A further strength is that we have based the assessment on two validated outcome measures, that are widely used for patients with chronic low back pain [23], and a cost parameter that was informed from national registries with no missing values.

A methodological weakness of the study is that SF-6D and EQ-5D measure different aspects of health-related quality of life. The EQ-5D has shown to lead to greater change over time compared to SF-6D [24]. Nevertheless, we have no reason to believe that using the alternative instrument of the EQ-5D would impact the randomization groups differently and, accordingly, our conclusion would be unaffected.

As a weakness it should be discussed whether the societal perspective was fully informed. It might have been relevant to include patients' out-of-pocket costs such as over-the-counter pharmaceuticals, complementary medicine, informal care and like but due to this being relatively demanding in terms of patient compliance over a long period of follow-up, where regular measurements are needed to prevent recall bias, this has not been prioritized. In a recently published another study from Denmark this was attempted, but the patient-reported costs had to be excluded from the main analysis due to high rates of non-response [25].

Another weakness is non-response on the outcome parameters. Rather than ignoring non-response, we chose to use a simple imputation strategy although it could be argued that more complex models would more appropriately take the characteristics of non-responders into account. Our choice of the simpler strategy was made in favor of transparency and due to the relatively small scale of the sample size where large variation was observed even amongst responders. It is likely that we have underestimated non-responders' outcomes, but based on the non-response analysis we argue that such underestimation would be equally distributed between randomization groups.

External validity

The study was initiated in 2003 and inclusion of patients took place until 2008. It is likely that the surgical techniques as technologies have matured since the trial period, and the conclusion in principle applies for the techniques as they were in 2008. Another relevant disclaimer in terms of external validity is the item costs that are rarely generalizable across settings. Nevertheless, as the study conclusion is based on differences between techniques it is very unlikely that the direction of difference will be affected from higher or lower item costs and thus in this particular study, generalizability of item costs is not a key issue.

Conclusion

In this study, we have demonstrated that there are no significant differences in either cost or effects between TLIF and instrumented PLF from a Danish, societal perspective, based on 2 years of follow-up. In the light of this and previous results that have altogether not been able to demonstrate a convincing advantage of TLIF, we conclude that TLIF does not seem to be a relevant alternative to PLF from a societal point of view.

Acknowledgments The authors would like to thank Jakob Hjort, data manager at Public Health and Quality Improvement, central Denmark Region, for letting us use his Stata syntax for counting weeks of inability to participate in the labor force. This study was funded by a grant from the Danish Strategic Research Council for the project CESpine, with the objective of investigating cost-effectiveness in spine surgery.

Conflict of interest None declared.

References

1. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H,

Zanoli G (2006) European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 15(supplement 2):192–300
2. Høy K, Bünger C, Neidermann B, Helmig P, Hansen ES, Li H, Andersen T (2013) Transforaminal lumbar interbody fusion (TLIF) versus posterolateral instrumented fusion (PLF) in degenerative lumbardisorders: a randomized clinical trail with 2-year follow-up. Eur Spine J
3. Potter BK, Freeman BA, Verwiebe EG, Hall JM, Polly DW Jr, Kuklo TR (2005) Transforaminal lumbar interbody fusion: clinical and radiographic results and complications in 100 consecutive patients. J Spinal Disord Tech 18(4):337–346
4. Humphreys SC, Hodges SD, Patwardhan AG, Eck JC, Murphy RB, Covington LA (2001) Comparison of posterior and transforaminal approaches to lumbar interbody fusion. Spine (Phila Pa 1976). 26(5):567–571
5. Hee HT, Castro FP, Majd ME, Holt RT, Myers L (2001) Anterior/posterior lumbar fusion versus transforaminal lumbar interbody fusion: analysis of complications and predictive factors. J Spinal Disord 14(6):533–540
6. Soegaard R, Christensen FB (2006) Health economic evaluation in lumbar spinal fusion: a systematic literature review anno 2005. Eur Spine J 15:1165–1173
7. Søgaard R, Fritzell P (2011) Economic evaluation in lumbar spinal fusion. Eur Musculoskelet Rev 6(3):199–204
8. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL (2005) Methods for the economic evaluation of health care programmes. Oxford University Press, New York
9. Fenwick E, O'Brien BJ, Briggs A (2004) Cost-effectiveness acceptability curves—facts, fallacies and frequently asked questions. Health Econ 13(5):405–415
10. Statens Serum Institut. (2013) The National Health Insurance Service Register. Available from: http://www.ssi.dk/Sundhedsdataogit/Registre/Sygesikringsregister.aspx
11. Statens Serum Institut. (2013) The National Patient Register
12. Statens Serum Institut. (2012) Tariffs in 2012
13. Statistics Denmark (2013) Disposable income by region, gender, age and income range [cited 24th May 2013]. Available from: http://www.statistikbanken.dk/statbank5a/selectvarval/define.asp?PLanguage=0&subword=tabsel&MainTable=INDKP6&PXSId=146749&tablestyle=&ST=SD&buttons=0
14. Brazier J, Roberts J, Deverill M (2002) The estimation of preference-based measure of health from the SF-36. J Health Econ 2001(21):271–292
15. Fairbank JC, Pynsent PB (2000) The Oswestry Disability Index. Spine 25(22):2940–2952
16. Efron B, Tibshirani RJ (1993) An introduction to bootstrap, 2nd edn. Chapman and Hall, New York
17. Schofferman J, Slosar P, Reynolds J, Goldthwaite N, Koestler MA (2001) prospective randomized comparison of 270 degrees fusions to 360 degrees fusions(circumferential fusions). Spine (Phila Pa 1976) 26(10):E207–E212
18. Fritzell P, Hagg O, Jonsson D, Nordwall A (2004) Cost-effectiveness of lumbar fusion and nonsurgical treatment for chronic low back pain in the Swedish Lumbar Spine Study: a multicenter, randomized, controlled trial from the Swedish Lumbar Spine Study Group. Spine (Phila Pa 1976) 29(4):421–434
19. Rivero-Arias O, Campbell H, Gray A, Fairbank J, Frost H, Wilson-MacDonald J (2005) Surgical stabilisation of the spine compared with a programme of intensive rehabilitation of the management of patients with chronic low back pain: cost utility analysis based on a randomised controlled trial. Br Med J 330(7502):1239
20. Søgaard R, Bünger C, Christiansen T, Høy K, Eiskjaer SP, Christensen FB (2007) Circumferential fusion is dominant over posterolateral fusion in a long-term perspective: cost-utility evaluation of a randomized controlled trial in severe, chronic low back pain. Spine (Phila Pa 1976) 32(22):2405–2414
21. Freeman BJ, Steele NA, Sach TH, Hegarty J, Soegaard R (2007) ISSLS prize winner: cost-effectiveness of two forms of circumferential lumbar fusion: a prospective randomized controlled trial. Spine (Phila Pa 1976) 32(25):2891–2897
22. Nielsen PR, Andreasen J, Asmussen M, Tønnesen H (2008) Costs and quality of life for prehabilitation and early rehabilitation after surgery of the lumbar spine. BMC Health Serv Res 8:209
23. Johnson LG, Hellum C, Nygaard ØP, Storheim K, Brox JI, Rossvoll I, Leivseth G, Grotle M (2013) Comparison of the SF6D, the EQ5D, and the Oswestry Disability Index in patients with chronic low back pain and degenerative disc disease. BMC Musculoskelet Disord 14:148
24. Søgaard R, Christensen FB, Videbaek TS, Bünger C, Christiansen T (2009) Interchangeability of the EQ-5D and the SF-6D in long-lasting low back pain. Value Health 12(4):606–612
25. Oestergaard LG, Christensen FB, Nielsen CV, Bünger CE, Fruensgaard S, Søgaard R (2013) Early versus late initiation of rehabilitation after lumbar spinal fusion: economic evaluation alongside a RCT. Spine (Phila Pa 1976) 38(23):1979–1985

AARHUS
UNIVERSITY

Declaration of co-authorship

Full name of the PhD student: Kristian Winther Høy

This declaration concerns the following article/manuscript:

Title:	Transforaminal lumbar interbody fusion (TLIF) versus posterolateral instrumented fusion (PLF) in degenerative lumbar disorders: a randomized clinical trial with 2-year follow-up
Authors:	Kristian Høy, Cody Bünger, Bent Niederman, Peter Helmig, Ebbe Stender Hansen, Haisheng Li, Thomas Andersen

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: Kristian Høy, Cody Bünger, Bent Niederman, Peter Helmig, Ebbe Stender Hansen, Haisheng Li, Thomas Andersen. Transforaminal lumbar interbody fusion (TLIF) versus posterolateral instrumented fusion (PLF) in degenerative lumbar disorders: a randomized clinical trial with 2-year follow-up. Eur Spine J. 2013 Sep;22(9):2022-9. doi: 10.1007/s00586-013-2760-2. Epub 2013 Apr 13.PMID:23584162 Free PMC Article

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. No or little contribution
B. Has contributed (10-30 %)
C. Has contributed considerably (40-60 %)
D. Has done most of the work (70-90 %)
E. Has essentially done all the work

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	D
2. Planning of the experiments and methodology design and development	D
3. Involvement in the experimental work/clinical studies/data collection	D
4. Interpretation of the results	E
5. Writing of the first draft of the manuscript	E
6. Finalization of the manuscript and submission	E

Signatures of the co-authors

Date	Name	Signature
01.11.17	Kristian Høy	Kristian Høy

AARHUS
UNIVERSITY

01.11.17	Cody Bünger	
01.11.17	Bent Niederman	
01.11.17	Peter Helmig	
01.11.17	Ebbe Stender Hansen	
01.11.17	Haisheng Li	
01.11.17	Thomas Andersen	

In case of further co-authors please attach appendix

Date: 09.12.17

Signature of the PhD student

AARHUS
UNIVERSITY

Declaration of co-authorship

Full name of the PhD student: Kristian Winther Høy

This declaration concerns the following article/manuscript:

Title:	Addition of TLIF does not improve outcome over standard posterior instrumented fusion. 5-10 years long-term Follow-up: results from a RCT.
Authors:	Kristian Høy, Kamilla Troung, Thomas Andersen, Cody Bünger

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: Kristian Høy, Kamilla Troung, Thomas Andersen, Cody Bünger. Addition of TLIF does not improve outcome over standard posterior instrumented fusion. 5-10 years long-term Follow-up: results from a RCT. Eur Spine J. 2017 Mar; 26(3):658-665. doi: 10.1007/s00586-016-4592-3. Epub 2016 May 7. PMID: 27155825

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

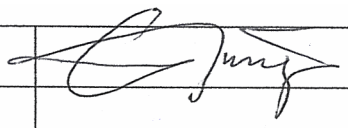
A. No or little contribution
B. Has contributed (10-30 %)
C. Has contributed considerably (40-60 %)
D. Has done most of the work (70-90 %)
E. Has essentially done all the work

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	D
2. Planning of the experiments and methodology design and development	D
3. Involvement in the experimental work/clinical studies/data collection	D
4. Interpretation of the results	E
5. Writing of the first draft of the manuscript	E
6. Finalization of the manuscript and submission	E

Signatures of the co-authors

Date	Name	Signature
01.11.17	Kristian Høy	Kristian Høy
01.11.17	Kamilla Troung	
01.11.17	Thomas Andersen	

AARHUS
UNIVERSITY

01.11.17	Cody Bünger	

In case of further co-authors please attach appendix

Date: 09.12.17

Christian Høy

Signature of the PhD student

AARHUS
UNIVERSITY

Declaration of co-authorship

Full name of the PhD student: Kristian Winther Høy

This declaration concerns the following article/manuscript:

Title:	Residual radiculopathy in TLIF surgery is not related to the side of cage insertion, but is higher in comparison to an instrumented standard posterolateral fusion measured by pain drawings. Results from a RCT
Authors:	Kristian Høy, Blazej Grycel, Thomas Andersen, Cody Bünger

The article/manuscript is: Published ☐ Accepted ☒ Submitted ☐ In preparation ☐

If published, state full reference:

If accepted or submitted, state journal: review Acta Orthopedica

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. No or little contribution
B. Has contributed (10-30 %)
C. Has contributed considerably (40-60 %)
D. Has done most of the work (70-90 %)
E. Has essentially done all the work

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	D
2. Planning of the experiments and methodology design and development	D
3. Involvement in the experimental work/clinical studies/data collection	D
4. Interpretation of the results	D
5. Writing of the first draft of the manuscript	E
6. Finalization of the manuscript and submission	E

Signatures of the co-authors

Date	Name	Signature
01.11.17	Kristian Høy	Kristian Høy
01.11.17	Blazej Grycel	Blazej Grycel
01.11.17	Thomas Andersen	
01.11.17	Cody Bünger	

AARHUS
UNIVERSITY

In case of further co-authors please attach appendix

Date: 09.12.2017

Christian Høy

Signature of the PhD student

AARHUS UNIVERSITY

Declaration of co-authorship

Full name of the PhD student: Kristian Winther Høy

This declaration concerns the following article/manuscript:

Title:	Transforaminal lumbar interbody fusion vs. posterolateral instrumented fusion - Cost-utility evaluation along side an RCT with 2-years of follow-up
Authors:	Ann Dement Christensen, Kristian Høy, Cody Bünger, Peter Helmig, Ebbe Stender Hansen, Thomas Andersen, Rikke Søgaard

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: Ann Dement Christensen, Kristian Høy, Cody Bünger, Peter Helmig, Ebbe Stender Hansen, Thomas Andersen, Rikke Søgaard. Transforaminal lumbar interbody fusion vs. posterolateral instrumented fusion - Cost-utility evaluation along side an RCT with 2-years of follow-up. Eur Spine J. 2014 May;23(5):1137-43. doi: 10.1007/s00586-014-3238-6. Epub 2014 Feb 21. PMID:24557326

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. No or little contribution
B. Has contributed (10-30 %)
C. Has contributed considerably (40-60 %)
D. Has done most of the work (70-90 %)
E. Has essentially done all the work

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	E
2. Planning of the experiments and methodology design and development	E
3. Involvement in the experimental work/clinical studies/data collection	C
4. Interpretation of the results	C
5. Writing of the first draft of the manuscript	B
6. Finalization of the manuscript and submission	C

Signatures of the co-authors

Date	Name	Signature
01.11.17	Ann Dement Christensen	
01.11.17	Kristian Høy	Kristian Høy

AARHUS
UNIVERSITY

01.11.17	Cody Bünger	
01.11.17	Peter Helmig	
01.11.17	Ebbe Stender Hansen	
01.11.17	Thomas Andersen	
01.11.17	Rikke Søgaard	

In case of further co-authors please attach appendix

Date:

Signature of the PhD student