

The effectiveness of decompression alone compared with additional fusion for lumbar spinal stenosis with degenerative spondylolisthesis: a pragmatic comparative non-inferiority observational study from the Norwegian Registry for Spine Surgery

Ivar M. Austevoll¹ · Rolf Gjestad^{4,5} · Jens Ivar Brox⁶ · Tore K. Solberg^{3,7} · Kjersti Storheim⁸ · Frode Rekeland¹ · Erland Hermansen^{1,2,9} · Kari Indrekvam^{1,2} · Christian Hellum^{3,10}

Received: 17 August 2015 / Revised: 2 June 2016 / Accepted: 22 June 2016 / Published online: 15 July 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract

Purpose To evaluate the effect of adding fusion to decompression in patients operated for lumbar spinal stenosis with a concomitant lumbar degenerative spondylolisthesis.

Methods After propensity score matching, 260 patients operated with decompression and fusion and 260 patients operated with decompression alone were compared. Primary outcome measures were leg and back pain [Numeric

Rating Scale (NRS), 0–10] and Oswestry Disability Index (ODI, 0–100) at 12 months.

Results At 12-month follow-up, the fusion group rated their pain significantly lower than the decompression alone group [leg pain 3.0 and 3.6, respectively, mean difference -0.6 , 95 % confidence interval (CI) -1.2 to -0.05 , $p = 0.03$ and back pain 3.3 and 3.9, respectively, mean difference -0.6 , 95 % CI -1.1 to -0.1 , $p = 0.02$]. ODI was not significantly different between the groups (21.0 versus 23.3, mean difference -2.3 , 95 % CI -5.8 to 1.1 , $p = 0.18$). Seventy-four percent of the fusion group and 63 % of the decompression alone group achieved a clinically important improvement in back pain (difference in proportion of responders = 11 %, 95 % CI 2–20 %, $p = 0.01$), corresponding to a number needed to treat of 9 patients (95 % CI 5–50). There was no significant difference in responder rate for leg pain (74 and 67 %, respectively, difference 7 %, 95 % CI -1 to 16 %, $p = 0.09$) or for ODI (67 and 59 %, respectively, difference 8 %, 95 % CI 0–18 %, $p = 0.06$). The duration of surgery and hospital stay was longer for the fusion group (mean difference 68 min, 95 % CI 58–78, $p < 0.01$ and mean difference 4.2 days, 95 % CI 3.5–4.8, $p < 0.01$).

Conclusion In the present non-inferiority study, we cannot conclude that decompression alone is as good as decompression with additional fusion. However, the small differences in the groups' effect sizes suggest that a considerable number of patients can be treated with decompression alone. A challenge in future studies will be to find the best treatment option for each patient.

Keywords Lumbar spinal stenosis · Degenerative spondylolisthesis · Treatment · Fusion

✉ Ivar M. Austevoll
imau@helse-bergen.no

¹ Kysthospitalet in Hagevik, Orthopedic Clinic, Haukeland University Hospital, Bergen, Norway

² Department of Clinical Medicine, University of Bergen, Bergen, Norway

³ Norwegian Registry for Spine Surgery (NORspine), University Hospital of Northern Norway, Tromsø, Norway

⁴ Division of Mental Health Care, Haukeland University Hospital, Bergen, Norway

⁵ Centre for Research and Education in Forensic Psychiatry, Haukeland University Hospital, Bergen, Norway

⁶ Department of Physical Medicine and Rehabilitation, Oslo University Hospital, Oslo, Norway

⁷ Department of Neurosurgery, University Hospital of Northern Norway, Tromsø, Norway

⁸ Communication and Research Unit for Musculoskeletal Disorders (FORMI), Oslo University Hospital, University of Oslo, Oslo, Norway

⁹ Department of Orthopedic Surgery, Ålesund Hospital, Møre and Romsdal Hospital Trust, Ålesund, Norway

¹⁰ Clinic for Surgery and Neurology, Department of Orthopedics, Oslo University Hospital, University of Oslo, Oslo, Norway

Background

Lumbar degenerative spondylolisthesis (LDS) is defined as an anterior slip of one vertebra relative to the adjacent one and is caused by degenerative changes in an intact vertebral arch [1]. Most patients have back pain, radiating pain in the lower extremities, and neurogenic claudication due to concomitant lumbar spinal stenosis [2].

The evidence for the effectiveness of surgical treatment of LDS is moderate [3–5], but the relative effectiveness and criteria for the appropriateness of the different surgical treatment options are unclear [6]. There are two dissimilar principal procedures used today: decompression alone and decompression in combination with a fusion procedure. Meta-analyses and systematic reviews conclude that there is moderate evidence of a tendency towards better outcome when fusion is added to decompression [1, 7–9].

In the US, both the rate of patients receiving instrumented fusion and the rate of complex fusion procedures have increased over the last 15 years [10, 11]. These procedures have higher costs and are associated with more severe complications than decompression alone [10, 11]. As well, more recently published studies provide some evidence against the use of extensive procedures when operating for LDS [12–14]. Hence, it is important to compare clinical outcome after decompression only with outcome after additional fusion procedures.

The purpose of the present observational cohort study was to compare the relative effectiveness of decompression alone and decompression with fusion in daily clinical practice, 3 and 12 months after surgery for LDS.

Materials and methods

We have followed the guidelines of the ‘Strengthening the Reporting of Observational Studies in Epidemiology’ (STROBE) statement [15].

Collection of data

All data were collected from the Norwegian Registry for Spine Surgery (NORSpine) [16]. NORSpine receives no funding from the industry, and 36 of the 40 centres which perform spinal surgery report to the registry. The patient forms consist of self-administered questionnaires completed before surgery (baseline) and at 3- and 12-month follow-up after surgery. The questionnaires contain patient-reported outcome measures and additional questions at baseline about demographics and lifestyle. During their hospital stays, patients’ data concerning comorbidity, radiological classifications, American Society of

Anesthesiologists (ASA) grade, perioperative complications, operation time and operation method were recorded by the surgeon. Peri- and postoperative complications and adverse events were reported by patients on the 3-month follow-up form.

Population/interventions

Inclusion criteria: (1) patients registered in NORSpine during the period 2007–2013; (2) patients assessed by the surgeon to have both spinal stenosis and degenerative spondylolisthesis; (3) patients operated with a decompression procedure or with a decompression and a fusion procedure.

Exclusion criteria: (1) former back surgery at the level to be operated on; (2) operated with an anterior approach; (3) operated with an interspinous device.

Based on the data from the surgeon forms, the patients were categorised as follows: the decompression group: only decompression was performed. The fusion group: the decompression was followed by an instrumented or a non-instrumented fusion procedure.

Primary outcome measures

1. Numeric Rating Scale (NRS 0–10) for leg and back pain.
2. Oswestry Disability Index (ODI 0–100) V. 2.0. ODI has been found to be an appropriate questionnaire for assessing treatment outcome in patients with degenerative spondylolisthesis [1]. It has been translated and validated for Norwegian patients [17].

Minimal clinically important difference (MCID) is estimated by Copay et al. to be 12.8 points for ODI, 1.2 for back pain and 1.6 for leg pain [18].

Secondary outcome measures

- (1) Global Perceived Effect (GPE) scale [19]: patient-rated satisfaction with treatment outcome was assessed using a single question on a seven-point descriptive scale with the answers ‘completely recovered’, ‘much improved’, ‘slightly improved’, ‘unchanged’, ‘slightly worse’, ‘much worse’ and ‘worse than ever’. The patients were trichotomised into ‘substantially improved’ (‘completely recovered’ and ‘much improved’), ‘slight or no change’ (‘slightly improved’, ‘unchanged’ and ‘slightly worse’) and ‘substantially deteriorated’ (‘much worse’ and ‘worse than ever’);
- (2) duration of surgery;
- (3) length of hospital stay;
- (4) perioperative complications and adverse events registered on the surgeon form;
- (5) patient-reported complications and adverse events.

Statistical methods

Propensity score matching (PSM) was used to make the distribution of observed baseline patient characteristics in the decompression and fusion groups as similar as possible [20]. Baseline covariates that could potentially bias the allocation to treatment were considered to be: age, gender, ASA grade, body mass index, smoking, ODI, NRS back pain, NRS leg pain, EQ5D, the presence of foraminal stenosis, degenerative disc disease, scoliosis, predominating back pain, number of levels operated on and neurological palsy. Some data were missing in some of these baseline covariates, resulting in varying sample sizes (N varying from 556 to 616). The net sample with intact information on all variables was $N = 464$. To prevent the loss of cases used for estimation, the multiple imputation method was used to impute information in covariates prior to calculation of the propensity score. The covariates scheduled for the propensity score were used as predictors for the imputation model. The variables were constrained in order to maintain the measurements' original range. After maximum 50 case draws, one value was imputed when data was missing [21]. The PSM scores were derived from a logistic regression model, and reflect a patient's theoretical baseline probability for being fused. Using the technique of '1:1 optimal matching without replacement', pairs of fused and non-fused patients with a difference in propensity scores less than 0.2 in logit of the standard deviation were formed [20].

In a previous study from NORSpine no differences in outcome measures were found when comparing compliers and non-compliers at follow-up [16]. We therefore assumed data to be missing at random [21]. The longitudinal difference analyses were based on the listwise deletion method. Analysis of mean group differences with 95 % confidence intervals (95 % CI) and calculation of proportion of responders at 12-month follow-up were performed. A responder was defined as a patient with an improvement in a primary outcome exceeding the MCID.

In order to include data from all three time points and hence make use of all available data, latent growth curve (LGC) models with full information maximum likelihood were used [22]. Thus, information from patients was not excluded if data were missing at one of the follow-ups. A piecewise growth model was estimated, including one linear change from baseline to 3 months and another from 3 to 12 months. It was thereby possible to make predictions concerning two different change periods: the change during the first 3 months and the change during the last 9 months. Two LGC models were estimated; one unconditional, with level and changes for both groups together, and one including the estimation of the group differences.

SPSS version 22 was used for descriptive statistics, tests for distribution of data, cross-tabulations with χ^2 test, Student t tests and Mann–Whitney U tests of group differences and linear and logistic regression modelling. Mplus 7.3 was used for analysing LGC models [23].

The sample size was based on the same assumption as an ongoing randomised multicentre trial on the same patient group (ClinicalTrials.gov Identifier: NCT02051374). In a non-inferiority design, with one sided testing, we wanted to detect whether the proportion of responders was more than 15 % higher in the fusion group than in the decompression group. This was to be tested by forming a 95 % confidence interval (CI) for the difference of proportions (i.e., percentage of responders in the fusion group minus percentage of responders in the decompression group) and was to be rejected if the upper limit of the CI was less than 15 %. If rejected we had to conclude that decompression alone was non-inferior, i.e., as good as, decompression with fusion. A 15 % difference in responder rate corresponds to a number needed to treat of seven patients ($NNT = 100/15 = 6.67$) [24]. Choosing a type 1 error = 0.05 and power = 0.80 gives a sample size of 116. Considering these assumptions, and adding 25 % for possible dropouts, a total of 155 patients were required in each group [25].

Results

Out of the 721 eligible patients, 616 met the inclusion criteria. Of these, 294 (48 %) had undergone a fusion procedure in addition to decompression. A total of 260 propensity score matched patients from each group were compared. Seventy-three percent returned the forms at 3 months, 85 % at 12 months and 94 % had at least one follow-up registration (Fig. 1). Table 1 presents baseline data before and after propensity score matching. Twenty-four percent of the decompression group and 66 % of the fusion group underwent a laminectomy. Otherwise a mid-line preserving decompression was performed. In the fusion group, 83 (32 %) patients were operated without use of instrumentation, 129 (50 %) with use of posterolateral pedicle screw instrumentation and 48 (18 %) with pedicle screws in combination with an interbody fusion.

Primary outcome measures

Figure 2 illustrates the changes in primary outcomes for both groups. Table 2 presents the measures at baseline and 3 and 12-month follow-up, the change in scores from baseline to 12 months and the proportion of patients with an improvement in a primary outcome exceeding the

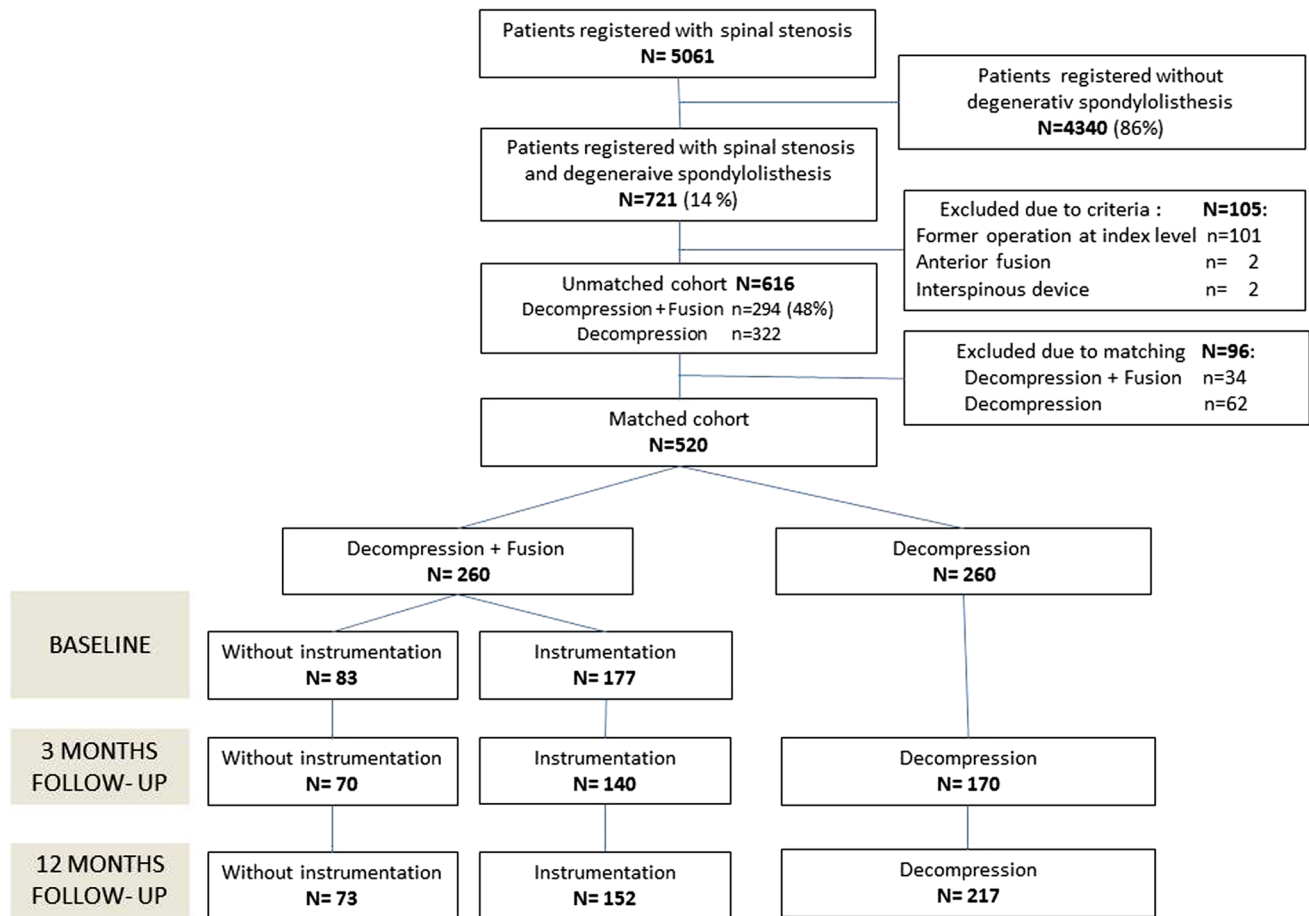


Fig. 1 Flow chart with enrolment and follow up

MCID. At 3- and 12-month follow-up, there were statistically significant lower levels of back and leg pain in the fusion group as compared with the decompression only group (mean difference at 12 months -0.6 , 95 % CI -1.2 to -0.05 , $p = 0.03$ for leg pain and mean difference -0.6 , 95 % CI -1.1 to -0.1 , $p = 0.02$ for back pain). There were no significant ODI differences between the groups when measured at both 3 and 12 months (mean difference at 12 months -2.3 , 95 % CI -5.8 to 1.1 , $p = 0.18$). At 12-month follow-up, we found a statistically significant difference between the proportion of patients in each group with a reduction of back pain exceeding MCID (74 versus 63 %, difference 11 %, 95 % CI 2–20 %, $p = 0.01$) favouring the fusion group. The corresponding difference in responder rate (fusion as compared to decompression alone) was not statistically significant for leg pain (74 and 67 %, respectively, difference 7 %, 95 % CI -1 to 16 %, $p = 0.09$) or for ODI (67 and 59 %, respectively, difference 8 %, 95 % CI 0–18 %, $p = 0.06$).

Table 3 shows the results analysed by LGC models. When both groups were analysed together we found a significant ($p < 0.01$) reduction in ODI and back and leg

pain from baseline to 3-month follow-up. For leg pain a significant deterioration was found from 3 to 12 months. At 12-month follow-up a significantly lower level of leg and back pain was noted in the fusion group as compared to the decompression alone group. Estimated differences were -0.55 for leg pain and -0.61 for back pain. We found no significant differences between the groups assessed by change in ODI and back and leg pain from either baseline to 3 months or from 3 to 12 months.

Secondary outcome measures

The mean duration of surgery was 103 min for the decompression group compared to 170 for the fusion group (mean difference 68 min, 95 % CI 58–78, $p < 0.01$). The hospital stay was 7.1 days for the fusion group and 2.9 days for the decompression group (mean difference 4.2 days, 95 % CI 3.5–4.8, $p < 0.01$). There were no significant differences in the proportion of perioperative complications [8 % (fusion) versus 5 %, $p = 0.29$] and patient-reported complications during the first 3 months [13 % (fusion) versus 15 %, $p = 0.53$] (Table 4).

Table 1 Baseline characteristics before and after propensity score matching

	Unmatched cohort (<i>n</i> = 616)			Propensity score matched cohort (<i>n</i> = 520)		
	Decompression (<i>n</i> = 322)	Decompression + fusion (<i>n</i> = 294)	<i>p</i> value	Decompression (<i>n</i> = 260)	Decompression + fusion (<i>n</i> = 260)	<i>p</i> value
Age; mean (SD)	67.8 (9.9)	65.8 (10.0)	0.01	66.7 (10.0)	66.3 (9.6)	0.67
Female	0.66	0.78	<0.01	0.72	0.75	0.55
High level of education (\geq Master)	0.28	0.30	0.51	0.27	0.29	0.63
ASA level (1–4); mean (SD)	2.04 (0.58)	2.01 (0.50)	0.50	2.00 (0.57)	2.02 (0.50)	0.74
Body mass index; mean (SD)	26.6 (4.6)	27.2 (5.2)	0.12	26.7 (4.6)	26.8 (4.6)	0.73
Smoker	0.18	0.19	0.82	0.19	0.20	0.91
Anxiety or depression according to EQ-5D	0.37	0.41	0.28	0.41	0.40	0.72
Disc degeneration	0.27	0.19	0.03	0.21	0.21	0.91
Degenerative scoliosis	0.07	0.08	0.64	0.06	0.06	0.72
Foraminal stenosis	0.11	0.11	0.06	0.09	0.10	0.88
Leg pain >1 year	0.70	0.71	0.76	0.70	0.72	0.78
Back pain >1 year	0.81	0.86	0.10	0.83	0.86	0.34
Daily use of analgesics	0.78	0.85	0.04	0.82	0.84	0.52
Neurological palsy	0.10	0.04	<0.01	0.03	0.04	0.82
Predominant back pain	0.25	0.31	0.13	0.26	0.27	0.69
ODI; mean (SD)	41.1 (15.5)	41.4 (14.7)	0.80	40.7 (15.0)	40.9 (14.5)	0.84
NRS leg pain; mean (SD)	6.8 (2.1)	6.6 (2.2)	0.22	6.8 (2.2)	6.7 (2.2)	0.52
NRS back pain; mean (SD)	6.8 (2.2)	6.9 (1.9)	0.51	6.9 (2.1)	6.8 (1.9)	0.88

Proportions are given unless otherwise stated

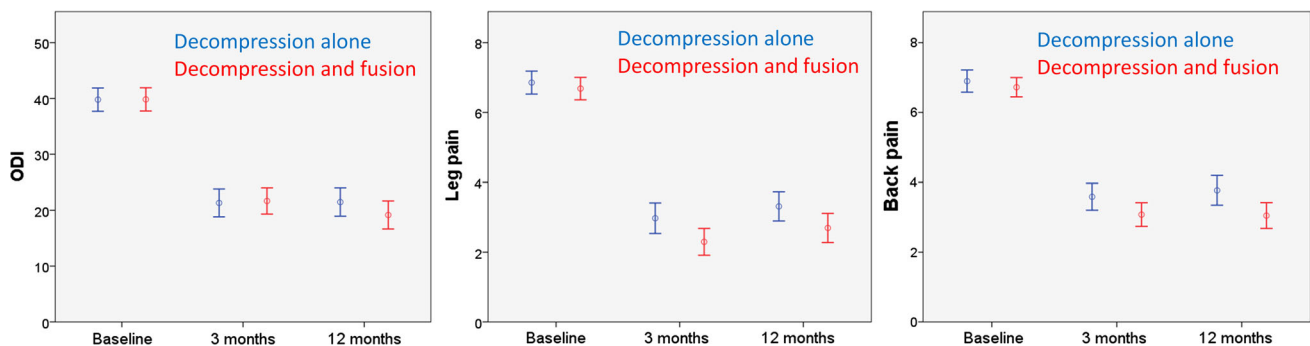


Fig. 2 Error bars for the propensity score matched cohort representing means and 95 % confidence interval for ODI, NRS back pain and NRS leg pain at baseline, 3 and 12 months

Table 5 presents the results for Global Perceived Effect score. Seventy percent of the fusion group versus 62 % of the decompression group rated themselves as ‘completely recovered’ or ‘much improved’ (difference 8 %, 95 % CI –11 to 17 %, $p = 0.09$), and 4 % in both groups rated themselves as ‘much worse’ or ‘worse than ever’ ($p = 0.96$).

Discussion

The difference in proportion of responders assessed by back pain was 11 %. This corresponds to a number needed to treat of 9 (95 % CI 5–50). Consequently, nine patients need to be fused to achieve one extra patient with reduced back pain exceeding MCID. The difference in responder

Table 2 Outcome measures for the propensity score matched cohort at baseline, 3 and 12 months, mean changes from baseline to 12 months and proportion of patients with improvement exceeding MCID from baseline to 12 months

	Decompression			Decompression + fusion			Mean diff (95 % CI)	<i>p</i> value
	Mean	SD	<i>N</i>	Mean	SD	<i>N</i>		
ODI								
Baseline	40.7	15.0	260	40.9	14.5	260		0.84
3 months	21.6	17.3	211	21.7	16.4	202	0.2 (−3.4 to 3.1)	0.92
12 months	23.3	18.5	218	21.0	18.2	224	−2.3 (−5.8 to 1.1)	0.18
Change 0–12 months	−17.5	16.1	218	−19.7	18.3	224	−2.2 (−5.5 to 1.0)	0.17
Proportion with change > MCID	0.59	0.49	218	0.67	0.50	224	0.09 (0.0 to 0.18)	0.06
NRS leg pain								
Baseline	6.8	2.2	260	6.7	2.2	260		0.52
3 months	3.0	3.0	210	2.4	2.6	198	−0.60 (−1.1 to −0.06)	0.03
12 months	3.6	2.9	212	3.0	3.0	215	−0.60 (−1.2 to −0.05)	0.03
Change 0–12 months	−3.3	3.1	212	−3.7	3.2	215	−0.46 (−1.1 to 0.15)	0.14
Proportion with change > MCID	0.67	0.47	212	0.74	0.43	215	0.07 (−0.01 to 0.16)	0.09
NRS back pain								
Baseline	6.9	2.1	260	6.8	1.9	260		0.88
3 months	3.6	2.7	212	3.1	2.3	202	−0.50 (−1.0 to −0.02)	0.04
12 months	3.9	2.9	216	3.3	2.6	223	−0.62 (−1.1 to −0.10)	0.02
Change 0–12 months	−3.0	2.9	216	−3.5	2.8	223	−0.56 (−1.1 to −0.02)	0.04
Proportion with change > MCID	0.63	0.49	216	0.74	0.44	223	0.11 (0.02 to 0.20)	0.01

Table 3 Results for the propensity score matched cohort estimated with Latent Growth Curve models

C	ODI	Back pain	Leg pain
Change 0–3 months			
Both groups, mean (SD)	−18.13 (16.6)**	−3.43 (2.8)**	−3.95 (3.2)**
Group difference (<i>b</i>)	−0.12 ns	−0.42 ns	−0.48 ns
Change 3–12 months			
Both groups, mean (SD)	−0.56 (13.9) ns	0.18 (2.4) ns	0.47 (2.6)**
Group difference (<i>b</i>)	−2.48 ns	−0.17 ns	0.02 ns
Level at 12 months			
Both groups, mean (SD)	22.10 (18.4)	3.60 (2.8)	3.27 (2.9)
Group difference (<i>b</i>)	−2.34 ns	−0.61*	−0.55*

The models shows estimated change from baseline to 3 months, estimated change from 3 to 12 months and estimated mean outcome level at 12 months. The regression weight (*b*) indicates the difference between the decompression and fusion groups (negative values indicate greater improvement in the fusion group)

Ns not significant

Significance testing of means: * *p* < 0.05, ** *p* < 0.01

rate was not statistically significant for leg pain or ODI. The upper limits of the 95 % confidence intervals for differences in responder rate were 20 % (back pain), 18 % (ODI) and 16 % (leg pain) all of which exceed the proposed clinically relevant difference of 15 %. Due to these results we cannot assert non-inferiority for decompression alone.

The mean difference at 12-month follow-up in leg and back pain (−0.6 for both) in favour of the fusion group is

considered to be small and of uncertain clinical importance. For ODI, the difference of −2.3 units in favour of fusion was not statistically significant.

A borderline significant difference in the percentage of patients rating themselves as ‘much improved’ or ‘completely recovered’ (70 % for the fusion group and 62 % for the decompression group) was also noted.

From a health-economic perspective, the considerably longer duration of surgery and hospital stay contribute to

Table 4 Surgical parameters and data from the hospital stays of the propensity score matched cohort

	Decompression	Decompression + fusion	Mean diff.	95 % CI	<i>p</i> value
Number of level operated on	1.4	1.4			0.88
Operated on L2/L3	0.07	0.08			0.61
Operated on L3/L4	0.39	0.34			0.24
Operated on L4/L5	0.79	0.83			0.22
Operated on L5/S1	0.12	0.13			0.69
Laminectomy	0.24	0.66	0.42	0.34–0.50	<0.01
Posterolateral fusion without use of instrumentation		0.32 (<i>n</i> = 83)			
Posterolateral fusion with pedicle screws		0.50 (<i>n</i> = 129)			
Posterolateral interbody fusion		0.03 (<i>n</i> = 8)			
Transforaminal interbody fusion		0.15 (<i>n</i> = 40)			
Mean operation time (in min)	103	170	68	58–78	<0.01
Dural tears	0.04	0.06			0.42
Perioperative complication	0.05	0.08			0.29
Patient-reported complications in the first 3 months	0.15	0.13			0.53
Mean hospital stay (in days)	2.9	7.1	4.2	3.5–4.8	<0.01

Proportions are given unless otherwise stated

Table 5 Patient-reported benefits of operation at 12-month follow-up according to the global perceived effect score for the propensity score matched cohort

	Decompression (%)	Decompression + fusion (%)
Completely recovered	21	25
Much improved	41	45
Slightly improved	22	20
Unchanged	8	4
Slightly worse	5	3
Much worse	2	2
Worse than ever	1	1

more costs for the fusion group. There were no significant differences between the groups in complications or adverse events.

The evidence for fusion in guidelines and meta-analyses [1, 7, 8] is often based on Herkovitz et al.'s paper from 1991 [26], which is considered to be of low quality according to the GRADE system [27]. Fifty consecutive patients were assigned to decompression with or without non-instrumented fusion. Pain reduction was greater in the fusion group. There were 24 out of 25 good or excellent in the fusion group compared to 9 out of 25 in the decompression alone group.

This year two randomised controlled trials were published [28, 29]. Ghogawala et al. compared laminectomy alone with laminectomy plus instrumented fusion (66 patients included) at 2-, 3- and 4-year follow-up. They

found the physical health-related quality of life to be slightly better and more clinically meaningful in the fusion group [29]. Försth et al. randomised 68 patients to decompression alone and 67 to decompression and fusion. No statistical or clinical differences were found between the groups at 2-year follow-up. Interestingly, the mean ODI score in the Swedish study corresponds strongly with the present study at baseline as well as at follow-up.

Several registry studies have been published over the past few years. A study of 213 patients (56 operated on with decompression alone) from one centre participating in the Spine Tango Registry showed better patient-reported outcomes when instrumented fusion was added to decompression. In that study, 86 % of the fusion group versus 70 % of the decompression alone group ($p = 0.01$) were defined as having a good outcome. When controlling for gender, age, comorbidity, number of levels operated on and baseline outcome scores in a multivariate regression analysis, fusion was a significant unique predictor for a 'good' global outcome [30]. Conversely, in a study from the Swedish Registry for Spine Surgery (Swespine), 655 patients operated with decompression only were compared with 651 operated with additional fusion. After adjusting for age, gender, smoking habits, use of analgesics and baseline outcome variables, no differences were found in clinical outcomes at 2-year follow-up; back pain 3.5 versus 3.2 ($p = 0.17$), leg pain 3.5 versus 3.2 ($p = 0.12$) and ODI 27 versus 27 ($p = 0.93$). The Swedish group concluded that a less invasive procedure should be recommended for elderly patients [13]. In another recently published study from Swespine, 839 patients (245 with decompression

alone) were analysed in relation to preoperative pain patterns. At 1-year follow-up, after adjusting for age, gender, duration of symptoms, earlier surgery and smoking, the fused patients with predominant back pain had significantly greater reduction in pain and ODI than the non-fused patients. At the 2-year follow-up, no differences were found regardless of pain pattern [14].

Strengths and limitations

Although propensity score matching was applied to reduce the confounding effects of differences in observed baseline characteristics between the groups, unobserved and unknown factors may be dissimilar in the treatment groups. For example, radiological biomechanical factors such as the size of the slip and the degree of instability could lead surgeons to choose one type of surgery over the other. The gold standard for evaluating efficacy in clinical research is a randomised controlled trial (RCT). An ongoing Norwegian RCT will provide better evidence about the additional effect of fusion and eliminate selection bias. However, the present study evaluates how the treatments work in daily clinical practice. The pragmatic design using multicentre registry data ensures a high external validity [31] and complements the recently published RCTs [28, 29] as well as the ongoing Norwegian RCT on the same subject.

The data has been analysed using different statistical methods, and results for the primary outcomes were consistent between the analyses. The clinical outcomes were investigated by the use of appropriate [1] and validated [17] PROMs. The proportion of patients missing in both follow-ups was low, which suggests that it is unlikely that missing data created biased results.

The present study does not address the long-term effect of a fusion procedure. The non-fused patients may develop recurrent stenosis or progression of the slip and the fused patients may be more prone to adjacent level degeneration [32, 33]. However, evidence regarding the influence of surgical procedure on long-term clinical results is sparse and inconsistent [5, 12, 14, 34–37].

Conclusion

In the present study, where the decision-making process was affected by physicians' and patients' personal preferences and perceptions, we cannot give evidence that decompression alone is as good as decompression with additional fusion. However, the small differences in the groups' effect sizes suggest that a considerable number of patients can be treated with decompression alone. In the future, well-powered studies should focus on finding the

best individual treatment option for patients with degenerative spondylolisthesis.

Compliance with ethical standards

Funding Helse Bergen HF, Haukeland University Hospital, Bergen, Norway.

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

See Tables 6, 7 and 8.

Table 6 Baseline characteristic for patients excluded due to matching. Proportions are given unless stated otherwise

	Decompression alone		Decompression + fusion		P
	(N = 62)		(N=34)		
Age; Mean (SD)	72.3	(8.4)	61.5	(11.5)	<0.01
Female; Mean (SD)	0.42	(0.50)	1.0	(0)	<0.01
Foraminal stenosis	0.19	(0.39)	0.18	(0.39)	0.99
ASA grade; Mean (SD)	2.2	(0.60)	2.0	(0.46)	0.03
BMI; Mean (SD)	26.1	(4.6)	29.9	(8.1)	0.02
Smoking	0.13	(0.34)	0.12	(0.33)	0.15
NRS Leg pain baseline; Mean (SD)	7.1	(3.2)	6.0	(2.5)	0.03
NRS Back pain baseline; Mean (SD)	6.5	(2.4)	7.4	(1.9)	0.06
ODI baseline; Mean (SD)	42.7	(17.3)	44.8	(15.8)	0.57
EQ5D baseline	0.38	(0.31)	0.25	(0.32)	0.05
Disc degeneration	0.52	(0.50)	0.06	(0.23)	<0.01
Degenerative scoliosis	0.11	(0.32)	0.18	(0.39)	0.39
Predominating back pain	0.23	(0.42)	0.56	(0.50)	<0.01
Number of level operated on; Mean (SD)	1.4	(0.48)	2.1	(1.1)	<0.01
Neurological palsy	0.35	(0.48)	0.03	(0.17)	<0.01

Table 7 Proportion of fusion performed in each hospital

Hospital number	N	Fusion rate (proportion)
1	66	0
2	26	0.08
3	28	0.11
4	9	0.11
5	24	0.13
6	10	0.20
7	49	0.24

Table 7 continued

Hospital number	<i>N</i>	Fusion rate (proportion)
8	15	0.40
9	23	0.48
10	10	0.50
11	11	0.55
12	35	0.60
13	42	0.64
14	118	0.66
15	16	0.69
16	10	0.70
17	12	0.75
18	19	0.79
19	88	0.80
20	5	1.00
	616	Total

Total number of operated patients (*N*), and proportion (fusion rate) operated on with an additional fusion procedure (i.e., “Hospital 9” has operated 23 patients, 48 % of them with an additional fusion)

Table 8 Group statistics for decompression alone versus decompression and additional instrumented fusion. Proportions are given unless stated otherwise

	Decompression alone		Decompression + instrumented fusion		<i>P</i>
	(<i>N</i> =260)	(<i>N</i> =177)	(<i>N</i> =260)	(<i>N</i> =177)	
Age; Mean (SD)	66.7	(10.0)	64.6	(9.7)	0.03
Female	0.72	(0.45)	0.71	(0.46)	0.70
Body Mass Index; Mean (SD)	26.7	(4.6)	26.7	(4.4)	0.99
Smoking	0.19	(0.40)	0.20	(0.40)	0.79
Neurological palsy	0.03	0.18	0.04	0.19	0.78
ODI score baseline; Mean (SD)	40.7	(15.0)	41.3	(14.2)	0.67
NRS Leg pain baseline; Mean (SD)	6.8	(2.2)	6.8	(2.2)	0.91
NRS Back pain baseline; Mean (SD)	6.9	(2.1)	6.8	(2.1)	0.91
ODI score 3 months; Mean (SD)	21.6	(17.3)	22.0	(16.9)	0.39
NRS Leg pain 3 months; Mean (SD)	3.0	(3.0)	2.4	(2.7)	0.07
NRS Back pain 3 months; Mean (SD)	3.6	(2.7)	3.3	(2.3)	0.81
ODI score 12 months; Mean (SD)	23.3	(18.5)	22.3	(18.8)	0.59
NRS Leg pain 12 months; Mean (SD)	3.6	(0.42)	3.0	(2.9)	0.10
NRS Back pain 12 month; Mean (SD)	3.9	(2.9)	3.6	(2.9)	0.23
Proportion with Δ ODI > MCID	0.59		0.66		0.13
Proportion with Δ leg pain > MCID	0.74		0.67		0.12
Proportion with Δ back pain > MCID	0.63		0.72		0.07

References

- Watters WC 3rd, Bono CM, Gilbert TJ, Kreiner DS, Mazanec DJ, Shaffer WO, Baisden J, Easa JE, Fernand R, Ghiselli G, Heggeness MH, Mendel RC, O’Neill C, Reitman CA, Resnick DK, Summers JT, Timmons RB, Toton JF, North American Spine S (2009) An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spondylolisthesis. *Spine J* 9:609–614
- Cauchoix J, Benoist M, Chassaing V (1976) Degenerative spondylolisthesis. *Clin Orthop* 115:122–129
- Mardjetko SM, Connolly PJ, Shott S (1994) Degenerative lumbar spondylolisthesis. A meta-analysis of literature 1970–1993. *Spine* 19:2256s–2265s
- Weinstein JN, Lurie JD, Tosteson TD, Hanscom B, Tosteson AN, Blood EA, Birkmeyer NJ, Hilibrand AS, Herkowitz H, Cammisia FP, Albert TJ, Emery SE, Lenke LG, Abdu WA, Longley M, Errico TJ, Hu SS (2007) Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *NEJM* 356:2257–2270
- Weinstein JN, Lurie JD, Tosteson TD, Zhao W, Blood EA, Tosteson AN, Birkmeyer N, Herkowitz H, Longley M, Lenke L, Emery S, Hu SS (2009) Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. Four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *J Bone Jt Surg (Am)* 91:1295–1304
- Mannion AF, Pittet V, Steiger F, Vader JP, Becker HJ, Porchet F (2014) Development of appropriateness criteria for the surgical treatment of symptomatic lumbar degenerative spondylolisthesis (LDS). *Eur Spine J* 23:1903–1917
- Martin CR, Gruszczynski AT, Braunsfurth HA, Fallatah SM, O’Neil J, Wai EK (2007) The surgical management of degenerative lumbar spondylolisthesis: a systematic review. *Spine* 32:1791–1798
- Resnick DK, Watters WC 3rd, Sharan A, Mummaneni PV, Dailey AT, Wang JC, Choudhri TF, Eck J, Ghogawala Z, Groff MW, Dhall SS, Kaiser MG (2014) Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 9: lumbar fusion for stenosis with spondylolisthesis. *J Neurosurg Spine* 21:54–61
- Steiger F, Becker HJ, Standaert CJ, Balague F, Vader JP, Porchet F, Mannion AF (2014) Surgery in lumbar degenerative spondylolisthesis: indications, outcomes and complications. A systematic review. *Eur Spine J* 23:945–973
- Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG (2010) Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA* 303:1259–1265
- Kepler CK, Vaccaro AR, Hilibrand AS, Anderson DG, Rihn JA, Albert TJ, Radcliff KE (2014) National trends in the use of fusion techniques to treat degenerative spondylolisthesis. *Spine* 39:1584–1589
- Chang HS, Fujisawa N, Tsuchiya T, Oya S, Matsui T (2014) Degenerative spondylolisthesis does not affect the outcome of unilateral laminotomy with bilateral decompression in patients with lumbar stenosis. *Spine* 39:400–408
- Forsth P, Michaelsson K, Sanden B (2013) Does fusion improve the outcome after decompressive surgery for lumbar spinal stenosis?: a two-year follow-up study involving 5390 patients. *Bone Joint J* 95-B:960–965
- Sigmundsson FG, Jonsson B, Stromqvist B (2015) Outcome of decompression with and without fusion in spinal stenosis with

- degenerative spondylolisthesis in relation to preoperative pain pattern: a register study of 1624 patients. *Spine J* 15:638–646
15. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 370:1453–1457
 16. Solberg TK, Sorlie A, Sjaavik K, Nygaard OP, Ingebrigtsen T (2011) Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? *Acta Orthop* 82:56–63
 17. Grotle M, Brox JI, Vollestad NK (2003) Cross-cultural adaptation of the Norwegian versions of the Roland–Morris Disability Questionnaire and the Oswestry Disability Index. *J Rehabil Med* 35:241–247
 18. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY (2008) Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. *Spine J* 8:968–974
 19. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ (2010) Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 63:760–766
 20. Austin PC (2011) An introduction to Propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res* 46:399–424
 21. Schafer JL, Graham JW (2002) Missing data: our view of the state of the art. *Psychol Methods* 7:147–177
 22. Bollen KA, Curran PJ (2006) Latent curve models: a structural equation perspective. Wiley, Hoboken
 23. Muthén LK, Muthén BO (2014) Mplus 7.3. In: Muthén, Muthén, 3463 Stoner Avenue, CA 90066, Los Angeles
 24. Katz N, Paillard FC, Van Inwegen R (2015) A review of the use of the number needed to treat to evaluate the efficacy of analgesics. *J Pain* 16:116–123
 25. Blackwelder WC, Chang MA (1984) Sample size graphs for “proving the null hypothesis”. *Control Clin Trials* 5:97–105
 26. Herkowitz HN, Kurz LT (1991) Degenerative lumbar spondylolisthesis with spinal stenosis. A prospective study comparing decompression with decompression and intertransverse process arthrodesis. *J Bone Joint Surg (Am)* 73:802–808
 27. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O’Connell D, Oxman AD, Phillips B, Schunemann HJ, Edejer T, Varonen H, Vist GE, Williams JW Jr, Zaza S (2004) Grading quality of evidence and strength of recommendations. *BMJ* 328:1490
 28. Forsth P, Olafsson G, Carlsson T, Frost A, Borgstrom F, Fritzell P, Ohagen P, Michaelsson K, Sanden B (2016) A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. *NEJM* 374:1413–1423
 29. Ghogawala Z, Dziura J, Butler WE, Dai F, Terrin N, Magge SN, Coumans JV, Harrington JF, Amin-Hanjani S, Schwartz JS, Sonntag VK, Barker FG 2nd, Benzel EC (2016) Laminectomy plus fusion versus laminectomy alone for lumbar spondylolisthesis. *NEJM* 374:1424–1434
 30. Kleinstueck FS, Fekete TF, Mannion AF, Grob D, Porchet F, Mutter U, Jeszenszky D (2012) To fuse or not to fuse in lumbar degenerative spondylolisthesis: do baseline symptoms help provide the answer? *Eur Spine J* 21:268–275
 31. Inacio MC, Paxton EW, Dillon MT (2016) Understanding Orthopaedic Registry Studies: a comparison with clinical studies. *J Bone Jt Surg Am* 98:e3
 32. Ekman P, Moller H, Shalabi A, Yu YX, Hedlund R (2009) A prospective randomised study on the long-term effect of lumbar fusion on adjacent disc degeneration. *Eur Spine J* 18:1175–1186
 33. Mannion AF, Leivseth G, Brox JI, Fritzell P, Hagg O, Fairbank JC (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* 39:1373–1383
 34. Abdu WA, Lurie JD, Spratt KF, Tosteson AN, Zhao W, Tosteson TD, Herkowitz H, Longely M, Boden SD, Emery S, Weinstein JN (2009) Degenerative spondylolisthesis: does fusion method influence outcome? Four-year results of the spine patient outcomes research trial. *Spine* 34:2351–2360
 35. Fox MW, Onofrio BM, Onofrio BM, Hanssen AD (1996) Clinical outcomes and radiological instability following decompressive lumbar laminectomy for degenerative spinal stenosis: a comparison of patients undergoing concomitant arthrodesis versus decompression alone. *J Neurosurg* 85:793–802
 36. Kormblum MB, Fischgrund JS, Herkowitz HN, Abraham DA, Berkower DL, Ditkoff JS (2004) Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective long-term study comparing fusion and pseudarthrosis. *Spine* 29:726–733 (**discussion 733–724**)
 37. Tsutsumimoto T, Shimogata M, Yoshimura Y, Misawa H (2008) Union versus nonunion after posterolateral lumbar fusion: a comparison of long-term surgical outcomes in patients with degenerative lumbar spondylolisthesis. *Eur Spine J* 17:1107–1112