The Spine Journal 16 (2016) 588–590

Commentary

# 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

Anne F. Mannion, PhD[a,*], Jens-Ivar Brox, MD, PhD[b], Jeremy C. Fairbank, MD, FRCS[c]

[a]*Spine Centre Division, Department of Teaching, Research and Development, Schulthess Klinik, Lengghalde 2, 8008 Zürich, Switzerland*
[b]*Department of Physical Medicine and Rehabilitation, Oslo University Hospital, Kirkeveien 166, 0424 Oslo, Norway*
[c]*Nuffield Orthopaedic Centre, Windmill Road, Headington, Oxford OX3 7HE, UK*

Received 23 October 2015; accepted 4 December 2015

**COMMENTARY ON:** Hedlund R, Johansson C, Hagg O, Fritzell P, Tullberg T. Swedish Lumbar Spine Study Group. The long-term outcome of lumbar fusion in the Swedish lumbar spine study. Spine J 2016;16:579–87 (*in this issue*).

Chronic low back pain (LBP) represents an enormous challenge to our health-care systems, and trying to find the role, if any, for its surgical treatment is a major public health issue. Current guidelines advocate the use of multidisciplinary cognitive-behavioral and exercise rehabilitation programs as first-line treatments for chronic LBP, with fusion surgery being considered only if such non-operative treatments are unavailable or have failed to improve the condition [1]. The recently published long-term results of three randomized controlled trials (RCTs) carried out in the United Kingdom and Norway support these recommendations, having found no evidence for the superiority of surgery at the 11-year follow-up [2]. The long-term follow-up of the Swedish RCT is published in this edition [3].

## Comparable treatment groups across the RCTs?

A systematic review of the midterm results of the RCTs from Sweden, Norway, and the United Kingdom highlighted the fact that the nature of the surgical treatments and their midterm outcomes were comparable across the trials [4]. However, the non-operative group in the Swedish study was not considered to have received contemporary evidence-based conservative treatment, suggesting that it was the poor results in this comparator group that accounted for the apparent superiority of fusion in the Swedish study. The non-operative group might hence be considered to represent "natural history" or, worse still, a nocebo group. As well as comparing non-operative and surgical treatment, the secondary aim of the Swedish study was to compare three different surgical techniques with one another. As a result, the two main groups (surgery and no surgery) were asymmetrical in size, and patients had a 3:1 chance of being randomized to surgery. Having already received conventional physiotherapy for years and failed to improve with it, randomization to one of the three surgical interventions was likely what the patient hoped for; possibly, assignment to "unstructured physiotherapy" [3] created harm and anger, precipitating a negative outcome [5]. This might have explained the 2-year results of the Swedish group. Previously, the authors of the Swedish Lumbar Spine Study claimed that their trial differed from the trials done in the United Kingdom and Norway in that the Swedish centers only included patients they were convinced were "good candidates for surgery" [6]. Interestingly, however, baseline characteristics in the four trials did not differ in any important aspect that would serve to substantiate their claim [4] (and unpublished data in connection with Mannion et al. [7]).

## Group changers and dropouts

Trials of operative versus non-operative treatment are fraught with difficulties. The problems of numerous group changers and dropouts must be faced and dealt with. The intention to treat (ITT) principle, in which patients are analyzed in relation to the groups to which they were randomized [8], remains the best approach to the analysis of data from RCTs. It eliminates known or unknown prognostic factors from being

FDA device/drug status:

Author disclosures: ***AFM:*** Nothing to disclose. ***J-IB:*** Nothing to disclose. ***JCF:*** Nothing to disclose.

* Corresponding author. Schulthess Klinik, Lengghalde 2, 8008 Zürich, Switzerland. Tel.: (41)-44-385-7584; fax: (41)-44-385-7590.

*E-mail address*: anne.mannion@yahoo.com (A.F. Mannion)

http://dx.doi.org/10.1016/j.spinee.2015.12.001

associated with a given treatment [9]. ITT analyses may be supported by analyses in which cases are grouped as "per protocol" (those who underwent the treatment to which they were assigned and completed all the follow-ups) or "as treated" (the treatment actually received) or "worst case" (group crossovers considered as "failing" the treatment to which they were randomized). These subgroup analyses present their own problems. Patients moving from non-operative to surgical treatment can be tracked with relative ease, but not those having surgery, failing it, and then going on to have successful non-operative treatment. Any subgrouping based on group changes in one direction only and on the assertion that a group change from non-operative to surgery indicates failure of non-operative treatment but revision surgery does not count as failure of surgery will inevitably be biased.

## Global assessment or serially measured outcomes?

An important issue in RCTs concerns the interpretation of results when differences arise between the primary and secondary outcomes, or between prospective and retrospective assessments of change. Retrospective assessments of "global outcome" are popular as an aggregate measure of all aspects of outcome of importance to the patient [10,11]. However, investigators using such scales should first ensure their validity by exploring the scale's relationships with pre- and post-values for corresponding domain scores [12,13]. If these global measures are truly measuring "change," their values should correlate as strongly with preoperative as with follow-up values of the outcomes they purport to reflect (eg, disability, pain, and quality of life) [12]. In many studies, these conditions are not fulfilled [10,14], suggesting recall bias or motivational bias (patients undergoing more cumbersome treatment overestimate their improvement [15]). In our own long-term data from the Norwegian and United Kingdom RCTs, the patients' "global assessment" correlated considerably more strongly with the Oswestry Disability Index (ODI) scores at long-term follow-up (r=0.70) than with the ODI values at baseline (r=0.17). It seems unlikely that, 11 to 13 years after treatment, patients can remember with any clarity their preoperative state, in order to quantify subsequent change. The longer the duration of follow-up, the less likely the "global assessment" is to be a reliable measure of change [14]. The changes derived by comparing serial measures on validated outcome instruments surely give a more truthful reflection of improvement or deterioration over time, without the concerns of recall or motivational bias.

The Swedish group's long-term results [3] now complement those of the combined Norwegian and United Kingdom studies with respect to the important, prospectively measured outcomes concerning pain and disability. Of interest, in the Swedish study, the mean ODI scores for the non-operative group appear to have improved since the 2-year follow-up [16], whereas those of the surgical group have remained stable. This has resulted in no significant group differences at the long-term follow-up [3].

In the Swedish study, only the "global assessment" showed any statistically significant differences favoring the surgical group, and only in the "as treated" or "per protocol" analyses: More of the surgical than the non-operative patients reported at long-term follow-up that they were better or much better [3]. In the ITT, even global assessment was not significantly different between the treatment groups. In their 2001 paper, the Swedish group states "pain, disability, global self-rating by the patient, and back-to-work were used as primary outcome measures in the study" [16]. We believe their follow-up report should have given equal emphasis to all these primary outcomes, rather than focusing on one singular retrospective rating of global outcome.

In "as treated" and "per protocol" analyses of the Swedish long-term data, the two treatment groups appeared to start with a similar ODI score and end with a similar ODI at follow-up, yet the surgery group had much superior ratings of improvement on "global assessment" [3]. This means a given reduction on the prospectively measured ODI was perceived as a "greater improvement" in the fusion group. This does not support the notion of the global assessment being a valid indicator of change over time and suggests possible motivational bias [15].

## Deaths during the studies

The Swedish group was able to acquire more information than in our shared study [7] regarding patients who had died during follow-up. They were since able to establish that the deaths were "all unrelated to CLBP." In the United Kingdom and Norway combined study [2], we saw a higher rate of known deaths in the surgical group (10 of 242 patients) than in the non-operative group (1 of 231) (Fisher exact test p=.01). It would have been interesting to know whether the Swedish study found the same. The deaths per treatment group are not reported in the manuscript but are possibly important.

## Interpretation of the results

We encourage the reader to consider carefully the use of the only statistically significant primary outcome as the focus of the long-term results in the Swedish study [3]. Comprehensive analyses and detailed results for the serial measures of pain and disability are found in the tables of their paper. We believe that the long-term results for all primary outcome measures have been under-communicated, particularly in the abstract. The abstract should have summarized the results for all the (original) primary outcomes, and for "global assessment" also using the ITT analysis. We disagree with the conclusion that "from the patient's perspective, reflected by the GA, lumbar fusion surgery is a valid treatment option in CLBP" [3]. This is a highly biased and selective interpretation; the ODI, pain, and quality of life measures also represent the patient's perspective. The abstract should have stated "on the other hand, the other primary outcome measures suggested no differences," not simply that "secondary outcome

measures suggested that there was still substantial disability at long-term after both treatments."

It is normal practice to look at a range of outcome measures to ensure that the overall data tell a consistent story. In all analyses other than the ITT (which did indeed deliver consistent findings with all outcome measures), the prospectively rated measures of pain, disability, and quality of life told the same story, with only the global assessment delivering different findings. Normally, one might go with the majority, but the Swedish group instead decided to focus on the global assessment. We consider this highly biased reporting and hope that there was no conflict of interest in this group of surgeon investigators (see Mannion et al. [17]).

We welcome the publication of the Swedish group's long-term outcomes, with a commendable follow-up rate [3]. Their data were originally analyzed in combination with those from the Norwegian and United Kingdom RCTs, but were then unexpectedly withdrawn, just before publication. It should now be possible to pool the findings in a mixed model analysis of the original data, or in a meta-analysis, to deliver an even stronger, evidence-based message to the spine community.

## References


[1] Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006; 15(Suppl. 2):S192–300.

[2] Mannion AF, Brox JI, Fairbank JC. 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 2013;13:1438–48.

[3] Hedlund R, Johansson C, Hagg O, Fritzell P, Tullberg T, Swedish Lumbar Spine Study Group. The long-term outcome of lumbar fusion in the Swedish Lumbar Spine Study. Spine J 2015;doi:10.1016/j.spinee.2015.08.065.

[4] Mirza SK, Deyo RA. Systematic review of randomized trials comparing lumbar fusion surgery to nonoperative care for treatment of chronic back pain. Spine 2007;32:816–23.

[5] Mooney V. Point of view on Fritzell P et al. 2001 Volvo award winner in clinical studies: lumbar fusion versus nonsurgical treatment for chronic low back pain. Spine volume 26, number 23, pp 2521–2534. Spine 2001;26:2532–3.

[6] Berg S, Tullberg T. Letter to the editor regarding Mannion, Brox, Fairbank. 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 2014;14:1087.

[7] Mannion AF, Leivseth G, Brox JI, Fritzell P, Hagg O, Fairbank JC. ISSLS prize winner: long-term follow-up suggests spinal fusion is associated with increased adjacent segment disc degeneration but without influence on clinical outcome: results of a combined follow-up from 4 randomized controlled trials. Spine 2014;39:1373–83.

[8] National Institute for Health and Care Excellence (NICE). Glossary. Available at: http://www.nice.org.uk/glossary?letter=I, 2015.

[9] Bubbar VK, Kreder HJ. The intention-to-treat principle: a primer for the orthopaedic surgeon. J Bone Joint Surg Am 2006;88:2097–9.

[10] Hagg A, Fritzell P, Oden A, Nordwall A. Simplifying outcome measurement: evaluation of instruments for measuring outcome after fusion surgery for chronic low back pain. Spine 2002;27:1213–22.

[11] Havakeshian S, Mannion AF. Negative beliefs and psychological disturbance in spine surgery patients: a cause or consequence of a poor treatment outcome? Eur Spine J 2013;22:2827–35.

[12] Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. J Clin Epidemiol 2002;55:900–8.

[13] Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 2009;17:163–70.

[14] Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. 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 2010;63:760–6, e1.

[15] Aseltine RH Jr, Carlson KJ, Fowler FJ Jr, Barry MJ. Comparing prospective and retrospective measures of treatment outcomes. Med Care 1995;33(Suppl. 4):AS67–76.

[16] Fritzell P, Hägg O, Wessberg P, Nordwall A. 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 2001;26:2521–34.

[17] Mannion AF, Brox JI, Fairbank JC. Reply to letter to the editor from Berg S, Tullberg T. Letter to the editor regarding Mannion, Brox, Fairbank. 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 2014;14:1087. Spine J 2015;15:380–1.