

Willems PC, Staal JB, et al. Spinal fusion for chronic low back pain: systematic review on the accuracy of tests for patient selection. Spine J 2012;

Design: Systematic review of prognostic studies

PICOS:

- Patient population: patients undergoing spinal fusion with at least 3 months of low back pain without signs of nerve root impingement, spinal stenosis, instability, or deformity
 - o Exclusion criteria were fracture, infection, objective motor neurologic deficit, ankylosing spondylitis, neoplasm, scoliosis, or kyphosis
- Intervention: a positive result on a test intended to select patients likely to benefit from spinal fusion
- Comparison: a negative test result on the same test for patient selection for spinal fusion
- Outcomes: pain, improvement, work disability, back-specific disability, reported in such a way that a relevant clinical cutoff could be defined and dichotomized into success and failure for the fusion operation
- Study types: Studies with at least 20 patients of whom some had positive and some had negative results on the index test

Study search and selection:

- Databases were PubMed and EMBASE through November 2010
- Search terms for tests of patient selection were related to MRI, orthosis, provocative discography, facet joint blocks, and temporary external transpedicular fixation (TETF)
- Two authors independently screened studies for inclusion and checked reference lists for additional eligible studies
- Risk of bias was assessed by modifying a version of a checklist commonly used for diagnostic accuracy studies, substituting selected diagnostic accuracy items with related items for prognosis
 - o For example, an item relating to whether the reference standard (gold standard) for a diagnostic test correctly classifies the target condition was replaced by an item relating to whether the outcome of fusion was measured by a validated measure of acceptable quality
 - o Similarly, an item for diagnostic tests which specifies that a short time elapse between the performance of the index test and the gold standard test (in order to ensure that the target condition has not changed between the two tests) was eliminated, substituting an item requiring at least a 2 year period between spinal fusion and measurement of the validated outcome
- Two-by-two tables were constructed which cross-tabulated test results (positive and negative) with outcomes (success and failure)
 - o Sensitivity, specificity, predictive value, and likelihood ratios were calculated from the tables in order to evaluate whether the tests

accurately predicted whether a patient would benefit from fusion, and therefore whether the test was appropriate for selecting fusion patients

Results:

- 1143 studies were screened from the literature search, of which 22 full text articles were assessed for eligibility
 - o 6 of these were excluded because only patients with a positive test result were selected for fusion; 6 other studies were excluded because the outcomes were reported only as mean values, and could not be divided into successful and unsuccessful operations, leaving 10 studies for analysis
- Of these 10 studies, 3 concerned immobilization, 4 concerned provocative discography, and 3 concerned TETF; there were not studies of MRI or of facet joint blocks which met inclusion criteria
- Shortcomings were noted in several of the included studies; often, uninterpretable tests were not reported, no clear cutoff point was defined for positive and negative test results, and not all patients who were tested had spinal fusion
- Due to differences in how the patient selection tests were performed, pooling of results (meta-analysis) was not done, and the systematic review was descriptive only
- Pain was the only outcome for which the predictive value of discography could be assessed, since pain was the only outcome consistently incorporated into the studies
- For orthosis (canvas corset, fiberglass cast, pantaloon plaster cast), response to immobilization did not predict or rule out a good outcome after spinal fusion
 - o Both positive and negative likelihood ratios had confidence intervals which included the null value of 1; neither a positive or a negative response to an orthosis predicted whether the patient had a successful outcome from fusion
- For provocative discography, four studies were examined
 - o Only one of the four studies had likelihood ratios which were statistically significant; while it had a positive predictive value of 88%, the negative predictive value was only 48%; only half of patients who would not do well with fusion were identified by the test
- For TETF, three studies were examined; sensitivities were high (80%-93%), but specificities were low (20% to 47%)
 - o One study had a significant negative likelihood ratio (0.15), this study had a 93% sensitivity, but a poor specificity (47%)

Authors' conclusions:

- There was risk of bias in most of the selected studies, which precludes firm conclusions from their reported findings
 - o In all but three studies, a proportion of patients with negative results did not have fusion and were excluded from the analysis (verification bias)

- The review focused on a limited number of individual tests; this provides no information about the combined use of prognostic tests to guide decision making
- Psychosocial factors like workers' compensation and smoking were not incorporated into the analysis; these factors are relevant to decision making, since they are associated with treatment failure for fusion
- The criteria for judging study quality are adapted from established criteria for the quality of studies of diagnostic tests, but have not been established as criteria for prognostic studies
- Studies which reported outcomes in terms of mean differences (rather than outcomes dichotomized into success/failure) were not included in the analysis
 - o One of these studies, using pressure-controlled discography, reported no mean differences across the study sample
 - o Another study of facet joint blocks similarly did not show mean differences between test results and surgical outcomes
- The findings of the current review, however, show that currently used tests do not improve results of fusion by improving patient selection, making it hard to propose spinal fusion as a standard treatment for chronic LBP

Comments:

- The study results apply only to patients with chronic LBP without instability, deformity, stenosis, or nerve root impingement
- The authors suspect that not operating on patients with negative test results could create a risk of verification bias in the studies which did this
 - o It is likely that the "negative test" patients who were denied fusion were denied fusion for reasons that were not reported, but these reasons are more likely than not to predict fusion failure
 - o Therefore, the most likely effect of these exclusions would be to lower the estimate of specificity of the tests (if these patients had been operated on and failed, the estimate of true negatives would be revised upward)
 - o However, the three studies which operated on all patients did not show better specificity than those which did not operate on all patients
 - o Verification bias, while possible, is not clearly shown to be very great
- There were 4 studies of discography, 2 of which injected control discs and 2 which did not; it is not clearly shown that the controlled disc studies were very predictive of fusion results
- Although the quality of the evidence is low, it is reasonable to conclude that discography, orthosis, and TETF are poor predictors of the outcome of surgery in patients with chronic LBP who do not have instability, deformity, stenosis, or nerve root impingement

Assessment: Adequate for some evidence that discography is not a reliable predictor of the outcome of fusion in patients with chronic LBP, and is unproven as a criterion for patient selection for fusion