Spine

SPINE Volume 36, Number 21S, pp S96–S109

Fusion *Versus* Nonoperative Care for Chronic Low Back Pain

Do Psychological Factors Affect Outcomes?

Michael D. Daubs, MD,* Daniel C. Norvell, PhD,† Robert McGuire, MD,‡ Robert Molinari, MD,§ Jeffrey T. Hermsmeyer, BS,† Daryl R. Fourney, MD, FRCSC, FACS,‖ J. P. Wolinsky, MD,¶ and Darrel Brodke, MD*

Study Design. Systematic review.

Objective. The objectives of this systematic review were to determine whether fusion is superior to conservative management in certain psychological subpopulations and to determine the most common psychological screening tests and their ability to predict outcome after treatment in patients with chronic lower back pain.

Summary of Background Data. Many studies have documented the effects of various psychological disorders on outcomes in the treatment of lower back pain. The question of whether patients with certain psychological disorders would benefit more from conservative treatment than fusion is not clear. Furthermore, the most appropriate screening tools for assessing psychological factors in the presence of treatment decision making should be recommended.

Methods. Systematic review of the literature, focused on randomized controlled trials to assess the heterogeneity of treatment effect of psychological factors on the outcomes of fusion *versus* nonoperative care of the treatment of chronic low back pain. In the analysis of psychological screening tests, we searched for the most commonly reported questionnaires and those that had been shown to predict lower back pain treatment outcomes.

Results. Few studies comparing fusion to conservative management reported differences in outcome by the presence or absence of a psychological disorder. Among those that did, we observed the effect of fusion compared with conservative management was more favorable in patients without a personality disorder, neuroticism, or depression. The most commonly reported, validated psychological screening tests for lower back pain are the Beck Depression Inventory, the Fear Avoidance Belief Questionnaire, the Spielberger Trait Anxiety Inventory, the Zung Depression Scale, and the Distress Risk Assessment Method.

Conclusion. Psychological disorders affect chronic lower back pain treatment outcomes. Patients with a personality disorder appear to respond more favorably to conservative management and those without a personality disorder more favorably to fusion. Patients with higher depression and neuroticism scores may also respond more favorably to conservative management.

Clinical Recommendations. Recommendation 1: Chronic LBP patients with depression, neuroticism, and certain personality disorders should preferentially be treated nonoperatively. Strength of recommendation: Weak.

Recommendation 2: Consider the use of a validated psychological screening questionnaire such as the BDI, FABQ, DRAM, ZDI or STAI, when treating patients with CLBP. Strength of recommendation: Weak.

Key words: lower back pain, psychological, psychological screening tests, surgical outcome, systematic review. **Spine 2011;36:S96–S109**

The biopsychosocial model for medicine emphasizes the need for full consideration of all factors that may impact a patient's response to treatment, including biological, psychological, and medicolegal factors.[1] Treatment outcomes as measured by patient-reported outcome tools are not solely a result of the specific treatment methods utilized, but are also a reflection of the individual patients' perception of their treatment. This perception is influenced by many factors including the psychological issues that may be impacting the patient at the time of treatment. Although pain levels may be similar among patients, their perceived level of suffering may vary widely.

Several case series have reported the effects of psychological disorders on the treatment outcomes of lower back pain (LBP)[2–4] and most have shown that anxiety, depression, and somatic disorders have a deleterious effect on both operative and nonoperative treatment in individual case series. Unfortunately, without the evaluation of these subgroups in comparative studies, we are unable to determine whether

From the *Department of Orthopaedics, University of Utah, Salt Lake City, UT; †Spectrum Research, Inc., Tacoma, WA; ‡Department of Orthopedics and Rehabilitation, University of Mississippi Medical Center, Jackson, MS; §Department of Orthopaedics, University of Rochester, Rochester, NY; ‖Neurosurgery Residency Training Program Division of Neurosurgery, University of Saskatchewan Royal University Hospital Saskatchewan, Canada; ¶ Department of Neurosurgery, Johns Hopkins University, Baltimore, MD.

Acknowledgment date May 9, 2011. Revision date: July 8, 2011. Acceptance date: July 21, 2011.

The manuscript submitted does not contain information about medical device(s)/drug(s).

Professional Organization and Foundation funds were received to support this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence and reprint requests to Michael D Daubs, MD, University of Utah, Department of Orthopaedics, 590 Wakara Way, Salt Lake City, UT 84108; E-mail: michael.daubs@hsc.utah.edu.

DOI: 10.1097/BRS.0b013e31822ef6b9

certain psychological subgroups respond more favorably to fusion or conservative management in those patients where the best treatment is unknown. Such data would aid in the challenge of treatment decision making. Results of spinal surgery for chronic low back pain in randomized controlled trials (RCTs) are less than encouraging.[5] This may be in part a result of classifying chronic low back pain as a homogeneous entity when in fact it is heterogeneous.[6–8] Results from RCTs represent average effects (population means), and, while estimates of the average treatment effect are useful, some individuals will respond more positively (efficacy) or more negatively (safety) than the reported average. Such variation in results is termed heterogeneity of treatment effects (HTE).[9] One way to identify HTE is to analyze the effect of treatment in subgroups of patients with certain baseline characteristics. However, subgroup analyses are prone to spurious results due to the problem of multiple testing.[10] Many caution against subgroup analyses, especially *post hoc* comparisons.[11] Nevertheless, identification of subgroup effects in clinical trials can generate important hypotheses about potential factors that modify treatment effects. Given that only one treatment is evaluated in a case series, this design does not address the question of whether treatment differences vary according to differing subgroup characteristics.[12–15] Therefore, although we hypothesized that there would be few comparison studies that stratified findings by psychological factors, we felt it imperative to attempt to identify those that did in an effort to generate hypotheses and identify gaps for future research. In addition to the challenge of identifying subgroups that respond more favorably to fusion or conservative management, unfortunately there is no standardized tool for evaluating or screening patients for the various psychological disorders that are believed to have an effect on treatment outcomes. Without validated screening tools, physicians have been shown to have difficulties in properly recognizing psychological distress.[16,17] As a result, psychological issues have been infrequently measured or included in the analysis of lower back treatment outcomes, although most physicians would agree that it is important to psychologically evaluate LBP patients.[18]

When evaluating patients with low back pain, the question remains whether patients should have treatment decisions on the basis of the presence of any underlying psychological factors. In other words, should patients with certain psychological issues always be treated nonoperatively? And, if this is true, what is the best tool for detecting the psychological issues that may be the most deleterious to a good outcome?

The objective of our study was twofold: (1) to determine, through systematic review, whether or not certain psychological issues modify LBP patients' treatment outcomes and if so, whether these factors could be utilized to determine the best method of treatment (fusion *vs.* nonoperative) and (2) determine the most useful psychological screening tools available for the assessment of psychological disorders that most influence treatment outcomes.

MATERIALS AND METHODS

Electronic Literature Database

A systematic search was conducted in MEDLINE and the Cochrane Collaboration Library for literature published from 1990 through December 2010. We limited our results to humans and to articles published in the English language. Reference lists of key articles were also systematically checked. We hypothesized that the following potential psychological subgroups may modify the treatment effect for LBP: depression, stress/anxiety, and personality disorder. For our first objective, to evaluate whether the effects of treatment varied by *psychological* subgroups, we sought randomized controlled trials evaluating surgical fusion *versus* nonoperative management for chronic LBP. More specifically, we approached the literature to identify the following: (1) RCTs designed specifically for evaluating spine fusion *versus* conservative management stratifying the random assignment on one or more psychological subgroups, (2) RCTs designed specifically for evaluating spine fusion *versus* conservative management that included a subgroup analysis stratifying on one or more psychological subgroups, and (3) RCTs that compared spine fusion *versus* conservative management among patients within a specific psychological subgroup (*e.g.*, personality disorder) to compare with other RCTs that were conducted among patients without patients in this subgroup (*e.g.*, no personality disorder). We excluded studies that did not report treatment effects (*i.e.*, fusion *vs.* conservative management) separately for the subgroups being compared unless they performed a statistical test for determining if the subgroup modified the treatment effect (*i.e.*, test for interaction). For example, if the authors reported a multivariate regression that included a subgroup variable (*e.g.*, depression [yes/no]) and the treatment variable (*e.g.*, fusion/conservative management), without an interaction term, the study was excluded. We excluded studies comparing any surgery (as opposed to fusion specifically) to conservative management, surgery *versus* surgery, and case series (a series of patients all receiving the same treatment). Articles were also excluded if they were pediatric studies (<18 years of age), nonfusion surgeries, included patients with predominantly neurological involvement, predominantly spondylolisthesis or stenosis, tumor surgery, revision surgery, treatment for osteomyelitis, inflammatory arthritis, or trauma. Other exclusions included reviews, editorials, case reports, and non-English–written studies, and studies without subgroup analyses (Figure 1). For our second objective, our search process was divided into three key steps: (1) To identify the most common psychological screening tests reported in the low back pain literature, we identified studies by using the following search code: ([Psychological OR Depression OR "Mental health" OR "Psychiatric treatment"] AND ["Screening Tool" OR "Screening Instrument" OR "Risk Assessment"]) AND "Low Back Pain." From these studies, we compiled a list of all psychological screening tests that were cited in the literature. To evaluate the relative frequency of use of these common psychological screening tests,

	Inclusion	Exclusion
Patient	•Adults	•<18 years old
	•Chronic LBP •Centralized or radiating pain	•Predominant neurological involvement •Predominant spondylolisthesis or stenosis •Cancer, deformity, instability, infection, trauma
Prognostic factors	•Adverse mental health predictors: •Depression •Stress/anxiety •Personality disorders •Screening tests	
Outcome	•Pain •Physical function •Quality of life	•Cost effectiveness
Study Design	•Meta-analyses •RCTs •Comparative observational studies •Registry studies •Studies including subanalyses of risk factors	•No separate treatment effect for each subgroup of interest. •Included risk factor regression analysis, but did not do a test for interaction. •Case reports •Non-clinical studies •Case series

Figure 1. Inclusion and exclusion criteria.

we searched PubMed using the name of the screening test and the common abbreviation combined with the following term: "Low Back Pain." The search results were limited to human studies published in the English language with no date restriction. The titles and abstracts of the studies identified were checked to verify that the screening test of interest was reported. The total number of studies reporting on each screening test in the title or abstract was determined. For the screening tests with the highest frequency of citations (≥10), we searched for studies that evaluated their predictive validity for determining outcomes.

Data Extraction

Each retrieved citation was reviewed by two independently working reviewers (D.C.N. and E.E.). Some articles were excluded on the basis of information provided by the title or abstract if they clearly were not appropriate. Citations that appeared to be appropriate or those that could not be excluded unequivocally from the title and abstract were identified, and the corresponding full-text reports were reviewed by the two reviewers. Any disagreement between them was resolved by consensus. For our first objective, the following data were extracted for both the surgical fusion and conservatively managed groups if the data were available: outcome, risk factor or subpopulation, rates of outcome (where appropriate), pre- and or postoperative outcome scores, effect estimates (*e.g.*, odd ratio, relative risk, treatment effect), and associated *P* values. For our second objective, the following data were extracted and summarized for the most common psychological screening tests: name of the measure, frequency of citations in literature, description of measure, interpretation, population validated in, outcomes validated against, and results of the predictive validity evaluation.

Study Quality

For our first objective to identify psychological subgroups, level of evidence ratings were assigned to each article independently by two reviewers using criteria set by *The Journal of Bone and Joint Surgery, American Volume*[19] to delineate criteria associated with risk of bias and methodological quality described elsewhere.[20] Our second objective was descriptive in nature, and therefore, rating each individual article was not relevant.

Analysis

For our first objective, the focus of the analysis was to evaluate subgroups within larger trials. We performed all analyses on a study level. Outcome measures are reported based on the author's choice of measure for subgroup treatment effects. Data were not pooled since only one article meeting study criteria was identified. For rates of binary outcomes, we calculated risk differences (RD) and 95% confidence intervals between fusion and conservative management arms for the overall population and separately by subgroup. Risk differences are considered standardized effect estimates. The reporting of effect estimates facilitates the interpretation of the size of the effect of a specific treatment as opposed to the statistical significance. A forest plot of RDs with 95% confidence intervals was constructed comparing fusion to conservative management by subgroup to evaluate whether there was any heterogeneity in the treatment effect (*i.e.*, that a treatment worked better in some subgroups than others). Bold vertical lines represent the no effect point (at zero) and a dashed line represents the overall treatment effect level. For studies that reported continuous scores of a particular subgroup at baseline, we used paired *t* tests to compare the differences in baseline scores between fusion and conservative groups by outcome (*i.e.*, improved or not improved). Analyses were performed using Stata 9.0 (StataCorp LP, College Station, TX).[21]

For our second objective, all analyses were descriptive. We reported whether or not the most frequently cited screening tests had been successfully (or unsuccessfully) tested with respect to their predictive validity in a chronic LBP population as reported in the results from each manuscript.

Overall Strength of Body of Literature

For our first objective, we evaluated individual articles as described earlier. The initial strength of the overall body of evidence was considered "HIGH" if the majority of the studies were level I or II and "LOW" if the majority of the studies were level III or IV. We downgraded the body of evidence one or two levels on the basis of the following criteria: (1) inconsistency of results, (2) indirectness of evidence, (3) imprecision of the effect estimates (*e.g.*, wide confidence intervals), or (4) if the authors did not state *a priori* their plan to perform subgroup analyses and if there was no test for interaction. We upgraded the body of evidence one or two levels on the basis of the following criteria: (1) large magnitude of effect or (2) dose-response gradient. The overall strength of the body of literature was expressed in terms of our confidence in the estimate of effect and the impact that further research may have on the results. An overall strength of HIGH means we have high confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect. The overall strength of "MODERATE"

means we have moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate. A grade of LOW means we have low confidence that the evidence reflects the true effect. Further research is likely to change the confidence in the estimate of effect and likely to change the estimate. Finally, a grade of "INSUFFICIENT" means that evidence either is unavailable or does not permit a conclusion. A more detailed description of this process can be found in the "Methodology" section.[20] Our second objective was descriptive in nature, and therefore, rating the overall body of literature was not relevant.

RESULTS

Study Selection

For our first objective, we identified 127 total citations from our search strategy for our first objective. Of these, 93 were excluded by abstract and 34 full text articles were retrieved to determine if they met criteria. From these 34, 10 reported sub-group effects; however, only three reported treatment effects (fusion *vs*. conservative management) separately by a psychological subgroup. Two of these were excluded because they included patients with predominantly neurological involvement (Figure 2).

For our second objective, 45 studies were identified that reported the use of a psychological screening test for LBP. We selected the five most frequently sited screening tests (≥10 citations) to assess their predictive validity. We identified 18 studies evaluating the predictive validity of these tests.

Is Fusion Superior to Conservative Management in Certain Psychological Subpopulations?

Only one study was identified meeting our study criteria that compared outcomes by treatment group stratified by a psychological subgroup. This highlights the limitations of the literature comparing fusion to conservative management in psychosocial subgroups with chronic LBP and can only serve to provide hypotheses regarding the possibility of treatment effect heterogeneity by psychological factors. In the RCT by Hägg (n = 264 patients with severe chronic LBP) comparing fusion to nonsurgical care (Table 1), several psychological assessments were performed on patients undergoing both treatments to include personality traits (neuroticism, aggressiveness, social introversion, and impulsiveness) using the Karolinksa Scales of Personality, existence of a personality disorder (cluster A, B, or C) using the Swedish version of the Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders* (Third Edition Revised), and depressive symptoms using the Zung Depressive Scale. Eleven percent of fusion patients *with a personality disorder* were considered "improved" ("better" or "much better" using the Patient Global Assessment) *versus* 17% of conservatively managed patients 2 years after surgery (Table 2). In contrast, 18% of fusion patients *without a personality disorder* and 8% who were conservatively managed were considered "improved" 2 years after surgery. The risk difference comparing fusion to conservative management in those with a *personality disorder* was −6% (RD = −0.06; 95% CI, −0.39 to 0.27) in favor of conservative management. The risk difference for patients *without a personality disorder* was 10% (RD = 0.10; 95% CI, −0.003 to 0.21) in favor of fusion (Table 2). The authors did not report a test for interaction on these treatment effect differences, but in examining raw scores, those *with a personality disorder* benefited more from conservative management and those *without a personality disorder* benefited more from fusion; however, the confidence intervals overlapped likely because of the small sample size in the personality disorder group (Figure 3).

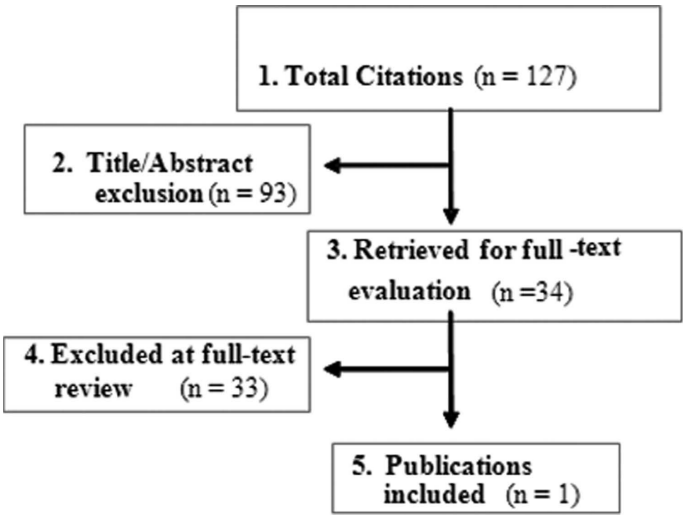

Figure 2. Flow chart showing results of literature search for psychological subgroups.

The same authors also reported baseline psychological measure scores for each treatment arm by those who "improved" and those who "did not improve" using the Patient Global Assessment as the outcome. The mean baseline depressive symptom score (Zung Depression Scale) (the higher the score the higher level of depression; major depression > 58) for those designated "improved" was 39.0 ± 13.4 points for the fusion group and 48.0 ± 11.3 points for the nonoperative group (Table 3). This difference was statistically significant (P = 0.009). The baseline differences for those "not improved" were similar between fusion and nonsurgical groups (Table 3). The mean difference comparing fusion to nonoperative groups in baseline depression scores among those who were designated "improved" was 9.0 points (95% CI: 2.3–5.7) and in those designated "not improved" 1.0 point (95% CI: −4.0 to 6.0) (Table 3). In other words, patients who improved with nonoperative treatment had higher levels of depression at baseline than those that improved with fusion. This may suggest that patients with higher levels of depression have better outcomes with nonoperative treatment.

There were similar results with the presence of a neurotic personality trait as determined by the Karolinksa Scales of Personality and translated by the authors as a person who is "tense and stiff, restless, uneasy, panicky, easily fatigued, remorseful, experiencing tremor and palpitations under stress." The mean difference comparing nonoperative

TABLE 1. Patient and Treatment Characteristics of Studies Reporting Treatment Effects Comparing Fusion to Conservative Management by Psychological Subgroups

Author (Year)	Study Design (LoE)	Follow-up (% Followed)	Demographics	Patient Characteristics	Interventions	Inclusion/Exclusion
Hägg (2003)	RCT multicenter Swedish lumbar spine study	2 years (90%)	*Surgery* n = 222, male: 50%, mean age: 43 years (25–64). *No surgery* n = 72, male: 49%, mean age: 44 years (26–63)	*Mean LBP duration* Surgery: 7.8 years (2–34) No surgery: 8.5 years (2–40) *Comorbidity* Surgery: 39.1% No surgery: 23.5% Smoking surgery: 40.6% No surgery: 49.3% Litigation/ compensation surgery: 60.4% No surgery: 64.5% Paid employment Surgery: 74% No surgery: 67%	Noninstrumented PLF (n = 73), instrumented PLF (n = 74), or instrumented PLIF (n = 75); all patients fused *in situ* with no intention of decompression; only segment L4–L5 and/or L5–S1 treated Physical therapy, supplemented with other forms of treatment such as education, pain relief (TENS, acupuncture, injections), cognitive and function training, and coping strategies.	*Inclusion* Aged 25–65 years Severe, chronic LBP of ≥ 2 years duration Back pain more pronounced that leg pain and no signs of nerve root compression Pain interpreted by surgeon as emanating from L4-L5 and/or L5-S1 with corresponding degenerative changes seen Must have been on sick leave for ≥ 1 year with failed conservative treatment Score of at least 7 of 10 for 10 questions reflecting function and working disability *Exclusion* Ongoing psychiatric illness Previous spine surgery other than successful removal of a herniate disc more than 2 years prior Spondylolisthesis, fractures, infection, inflammatory process, or neoplasm Painful and disabling arthritic hip joints and spinal stenosis

management to fusion in baseline neuroticism scores among those who were designated "improved" was 6.4 points (95% CI: 2.1–10.7) and in those designated "not improved" −0.9 points (95% CI: −4.5 to 2.7) (Table 3). This may suggest again that patients with higher baseline neuroticism scores benefit more from nonoperative management than fusion. There was little difference in baseline scores between nonoperative and fusion groups in those who did not improve.

TABLE 2. Study by Hägg Reporting Percent Improved Comparing Fusion to Conservative Management by Those With and Without a Personality Disorder at Baseline

			Fusion		Conservative			
			N (%)		N (%)			
Study	Outcome	Subgroup	A	B	A	B	Risk Difference	*P**
Hägg (2003)	Patient global assessment (% improved)	A: personality disorder B: no disorder	11% (n = 2/19)	18% (n = 21/117)	17% (n = 1/6)	8% (n = 3/40)	A: −0.06 (−0.39, .27) B: 0.10 (−0.003, .21)	NR

**Test for interaction.*

NR indicated not reported.

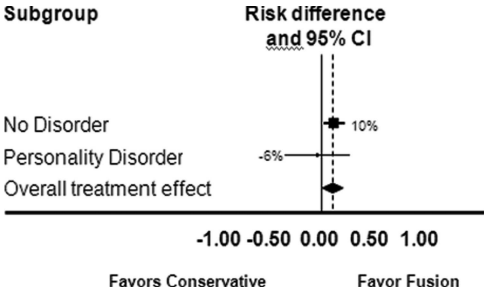

Figure 3. Forest plot representing the risk difference (RD) and 95% confidence interval comparing fusion to conservative management in the Personality Disorder and no Personality Disorder subgroups (and overall effect) in the study by Hägg.

What Are the Most Common Psychological Screening Tests and How Good Are They at Predicting Outcome After Treatment in Patients With Chronic LBP?

The following psychological screening tests were identified as the most commonly cited tests in the LBP literature: Beck Depression Inventory (BDI; n = 58), Fear Avoidance Belief Questionnaire (FABQ; n = 55), Coping Strategies Questionnaire (CSQ; n = 22), Zung Depression Scale (ZDS; n = 16), the Spielberger Trait Anxiety Inventory (STAI; n = 15), and the Distress and Risk Assessment Method (DRAM; n = 11) (Figure 4). All five instruments had at least one study evaluating its predictive validity in a LBP population. The following subsections provide a brief review of these findings in order of the most frequently cited in the literature. A summary of findings can be found in Table 4. Details of these studies are presented in Table 5.

Distress and Risk Assessment Method

This 45-item measure is a combination of the Modified ZDI (23 items) and the Modified Somatic Perceptions Questionnaire (22 items; only 13 are scored). The DRAM was proposed and evaluated by Main *et al.*[2] It separates patients into four categories (normal, at-risk, distressed-depressive, and distressed-somatic). In the study by Main *et al.* evaluating 98 patients with CLPB referred for orthopedic surgery, this measure was associated with pain, disability, and work status, 12 to 48 months after surgery. In another study of 66 patients undergoing lumbar discectomy, the DRAM was not predictive of Oswestry Disability Index (ODI) scores.[22] A final study of 102 patients undergoing lumbar spine surgery, the DRAM predicted work status change in leg or back pain, and the Dallas Pain Questionnaire 6 and 12 months after surgery.[3]

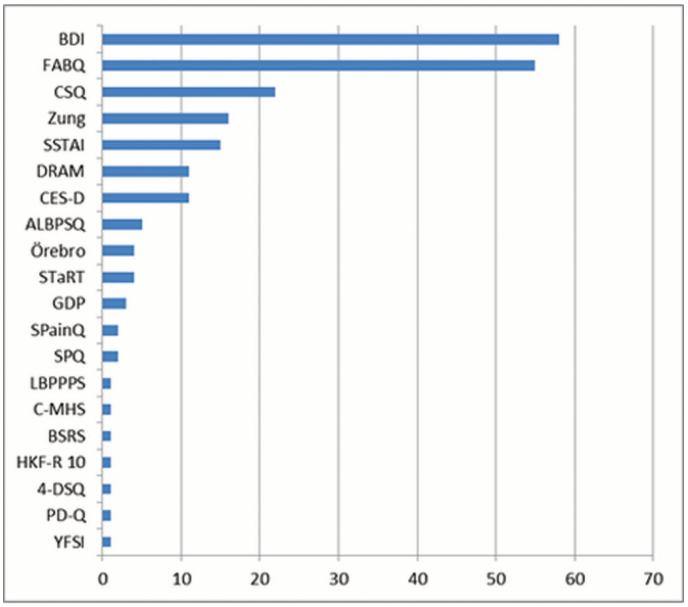

Figure 4. Frequency of citations for psychological screening tests in studies evaluating the treatment of chronic lower back pain.

TABLE 3. Study by Hägg Reporting Treatment Effects (Improved and Not Improved) Comparing Fusion to Conservative Management by Baseline Depression and Neuroticism Scores

Outcome*	Subgroup Scores	Treatment	Baseline Score (Points)		Mean Difference (Points)	95% CI
			Mean	SD		
Improved	Depression†	Conservative	48.0	11.3	9.0‡	2.3–15.7
		Fusion	39.0	13.4		
Not improved		Conservative	40.0	12.9	1.0	−4.0–6.0
		Fusion	39.0	13.3		
Improved	Neuroticism§	Conservative	56.5	8.8	6.4‡	2.1–10.7
		Fusion	50.1	8.3		
Not improved		Conservative	53.2	8.1	−0.9	−4.5–2.7
		Fusion	54.1	9.8		

**Patient Global Assessment (those who were "better" or "much better" were considered "improved" otherwise "not improved").*

†Zung Depressive Scale (higher the score the greater the depression).

‡Difference statistically significant subtracting mean baseline scores in conservative groups from fusion groups.

§Karolinska Scales of Personality.

TABLE 4. Summary of the Number of Items, Psychometric Properties,* Languages,† and Propriety‡ of the Most Common Psychosocial Measures

Measure	No. Items	Predictive Validity*	Languages†	Proprietary‡
Beck Depression Index (BDI)	21	X	X	NO
Fear Avoidance Belief Questionnaire	16	X	X	NO
Zung Depression Scale (ZDS)	20	X	X	NO
Distress and Risk Assessment Method (DRAM)	45	X		NO
State-Trait Anxiety Inventory (STAI)	40	X	X	NO

**An "X" indicates the measure was successfully evaluated for predictive validity.*

†An "X" indicates that the measure has been validated in a language(s) other than English.

‡An "X" indicates the measure is licensed or copyrighted and requires approval and fee to use. This status is subject to change and may not always be up to date.

Beck Depression Inventory

This 21-item self-report measure is used to measure severity of depression. The total score ranges from 0 to 63. Scores between 1 and 10 represent "normal ups and downs" and scores greater than 40 represent "extreme depression." Each increasing 10-point range represents a higher degree of depression. In a study evaluating 111 patients with acute radicular pain and a lumbar disc prolapse or protrusion that either had surgery (n = 73) or conservative treatment (n = 38), higher scores were found to be predictive of persistent chronic LBP.[23] The authors did not report differences by surgery or conservative management. Another study in 102 patients with symptomatic lumbar spinal stenosis, who underwent decompression, found higher preoperative BDI scores to be associated with 1-year postoperative functional ability (ODI), symptom severity, and poorer walking capacity.[24]

Fear-Avoidance Beliefs Questionnaire

This 16-item self-report measure has two subscales: physical activity and work. The physical activity score ranges from 0 to 24 and the work score ranges from 0 to 42; the higher the subscale scores, the greater the fear and avoidance beliefs. Four studies were identified evaluating the predictive validity of the FABQ in LBP populations. Populations were conservatively managed in three of the studies and one of the studies had a surgical and conservative arm. In one study, 42 subjects from a Middle Eastern culture with activity limiting LBP for more than 2 months enrolled in an exercise-based physical therapy program.[25] A multidimensional test battery was completed before and after a 10-week program of lumbar extensor muscle strengthening. An Arabic translation of the FABQ was used, which included a higher range of physical function scores. The association between the FABQ and clinically meaningful improvements in the Roland-Morris Disability Questionnaire (RMDQ) score was assessed. The physical activity subscale of the FABQ was associated with a negative outcome when the observed scores are ≥29. The work-specific subscale was not associated with clinically important improvements in the Roland Morris Disability Index (RMDI).

Another study investigated the FABQs ability to predict 6-month ODI scores for patients with LBP participating in physical therapy clinical trials.[26] Subjects (n = 160) were participants in 2 separate randomized trials investigating the efficacy of physical therapy interventions for LBP. The FABQ work scale was the better predictor of self-report of disability in this sample of patients participating in physical therapy clinical trials. Elevated physical activity scores were not associated with ODI scores.

Another study evaluated 108 patients with a 6-month history of mechanical low back pain, newly referred to an orthopedic outpatient clinic.[27] Subjects completed the FABQ along with an additional battery of instruments and then 6 months later completed the 36-Item Short Form Health Survey (physical composite score was reported) and the number of health care contacts during follow-up was recorded. These health care contacts included outpatient, inpatient, day case attendance, operations, and contact with any emergency services or health care professionals.

Higher scores for the FABQ were associated with impairment in subsequent physical health-related quality of life and number of health care contacts. The authors did not provide a distribution of patients who received surgical *versus* conservative management.

Finally, patients with acute (n = 123) and chronic (n = 50) LBP completed a comprehensive assessment, including the FABQ, were treated with physical therapy, and were followed at 3, 6, 9, and 12 months with a numeric pain rating scale and the ODI.[28] Patients with chronic LBP had more fear-avoidance beliefs for work than patients with acute LBP. Fear-avoidance beliefs predicted pain and disability at 12 months after adjusting for sociodemographic and pain variables.

Coping Strategies Questionnaire

This 48-item self-report questionnaire assesses seven different coping strategies. Patients rate the frequency with which they use the strategies on a 5-point scale. Two studies were identified assessing the predictive validity of this instrument in patients with LBP undergoing conservative management. The first study evaluated 84 subjects with chronic LBP undergoing physical therapy and found that the CSQ was not associated with a visual analog scale for pain or the RMDQ after controlling for the influence of catastrophic thinking and self-efficacy for pain control.[29] The second study evaluating 200 subjects with chronic LBP entering a Work Hardening Program also reported no significant

TABLE 5. Predictive Validity of Psychological Screening Tests in Studies Evaluating the Treatment of Chronic Low Back Pain

Instrument	Description	Interpretation	Population Tested	Outcome	Predictive Validity
Distress and Risk Assessment Method (DRAM)	A 45-item questionnaire made up of the following indices: Modified Zung Depression Index (23 items), Modified Somatic Perceptions Questionnaire (MSPQ) (22 items). Only 13 of the 22 items of the MSPQ are scored. Items scored on a 0 to 3 point scale.	Both indices scored separately. Type N (normal): Zung score <17. Type R (at risk): Zung score 17–33 and MSPQ score < 12. Type DD (distressed-depressive): Zung score > 33. Type DS (distressed-somatic): Zung score 17–33 and MSPQ > 12.	Patients with low back pain referred to orthopedic surgery clinic (N = 98) (41 years, 49% male)[2]	Pain Disability Work status	+
			Patients undergoing lumbar discectomy (N = 66) (38 years, 48.5% male)[22]	Oswestry Disability Index	-
			Patients undergoing lumbar spine surgery (N = 102) (47.3 ± 14.9, 51% male)[3]	Work status, change in leg or back pain, Dallas Pain Questionnaire	+
Beck Depression Inventory (BDI)	A 21-question self-report inventory used to measure the severity of depression. Items scored on a 0- to 3-point scale.	Maximum score: 63. Minimum score: 0. Total score: 1–10: normal ups and downs, 11–16: mild mood disturbance , 17–20: borderline clinical depression, 21–30: moderate depression, 31–40: severe depression, >40: extreme depression.	Patients with radicular pain and a lumbar disc prolapse or protrusion (N = 111) that either had surgery (n = 73) or conservative treatment (n = 38) (71.7 years, 61.3% male)[23]	Persistent chronic low back pain	+
			Patients with symptomatic lumbar spinal stenosis undergoing decompressive surgery (N = 102) (61.7 years, 42% male)[24]	Oswestry Disability Index, Stucki Questionnaire, Visual Analog Scale for pain, self-reported walking ability	+
			Patients with lumbar disc herniation treated with pharmacotherapy and/or physical therapy (N = 56) (19–21 years, 100% male)[31]	Modified Oswestry Disability Questionnaire	+
			Patients with acute low back pain < 6 months (n = 35) (50.2 ± 14.6 years, 94.3% male) or chronic low back pain of > 6 months (n = 75) (51.1 ± 11.5 years, 98.7% male) undergoing conservative treatment[33]	Pain	+

(*Continued*)

TABLE 5. (*Continued*)

Instrument	Description	Interpretation	Population Tested	Outcome	Predictive Validity
Fear-Avoidance Beliefs Questionnaire (FABQ)	Two fear-avoidance subscales (16 items): Physical activity Work Items 1, 8, 13, 14, and 16 included in scale, but not included in final score. Items included in final score are scored on a 0 to 6 point scale.	Maximum physical activity score: 24. Minimum physical activity score: 0. Maximum work score: 42. Minimum work score: 0. The higher the subscale scores, the greater the fear and avoidance beliefs.	Patients with 2 month history of activity limiting low back pain undergoing physical therapy (N = 42) (42.3 years, 52.3% male)[25]	Roland Morris Disability Questionnaire	+ (physical activity subscale) − (work subscale)
			Patients with 6-month history of mechanical low back pain newly referred to orthopedic outpatient clinic (N = 108) (39.9 ± 12.2 years, 55.6% male)[27]	SF-36 physical component, Client Socio-Demographic and Service Receipt Inventory (health care contacts)	+
			Patients with low back pain that have completed 4 weeks of physical therapy (N = 160) (34.6 years, 55.6% male)[26]	Oswestry Disability Index	+ (work subscale) - (physical activity subscale)
			Patients with acute low back pain of < 3 weeks (n = 123) (37.9 ± 10.1 years, 45% male) or chronic low back pain of > 3 months (n = 50) (40.4 ± 9.5 years, 38% male) undergoing physical therapy[28]	Numeric pain rating scale, Oswestry Disability Index	+
Coping Strategies Questionnaire (CSQ)	Assesses the following 7 different coping strategies (48 items): Diverting attention Ignoring pain sensations Reinterpreting pain sensations Coping self-statements Catastrophizing Praying or hoping Increasing activity level	Patients rate the frequency with which they use the strategies on a 5-point scale, from 1 ("I never do when in pain") to 5 (I very frequently do when in pain").	Patients with > 3 month history of chronic low back pain undergoing physical therapy (N = 84) (42 years, 55% male)[29]	Visual Analog Scale for pain, Roland-Morris Disability Questionnaire	-
			Patients with average 9 month history of low back pain assessed for entry into a Work Hardening Program (N = 200) (39 years, 70.5% male)[30]	Oswestry Disability Index, Symptom Checklist-90, return to work	-

(*Continued*)

TABLE 5. (*Continued*)

Instrument	Description	Interpretation	Population Tested	Outcome	Predictive Validity
Zung Depression Scale	A 20-question self-report scale used to measure the four common characteristics of depression: Pervasive effect Physiological equivalents Other disturbances Psychomotor activities There are 10 positively worded and 10 negatively worded questions. Items scored on a 1 to 4 point scale.	Maximum score: 80 Minimum score: 20 Total score: 20-49: Normal 50-59: Mildly depressed 60-69: Moderately depressed ≥70: Severely depressed	Volunteers with no history of serious low back pain (N = 403) (7.9% male, 27 years)[34]	First time back pain, Serious back pain	+
			Patients with > 3 month history of chronic low back pain (N = 102) (47.3 ± 14.9 years, 51% male) treated with lumbar fusion (n = 69), decompression (n = 30), or instrument removal (n = 3) [3]	Dallas Pain Questionnaire, work status, Numeric rating scale for back and leg pain	+
State-Trait Anxiety Inventory (STAI)	A 40-item scale providing separate measures of: State anxiety Trait anxiety State anxiety is a measure of the intensity of anxiety experienced at the time of assessment. Trait anxiety reflects the general tendency for experiencing anxiety. Items scored on a 1 to 4 point scale.	State and trait anxiety scored separately: Maximum score: 80 Minimum score: 20 The higher the score, the greater the anxiety.	Patients with acute low back pain < 6 months (n = 35) (50.2 ± 14.6 years, 94.3% male) or chronic low back pain of > 6 months (n = 75) (51.1 ± 11.5 years, 98.7% male) undergoing conservative treatment[33]	Pain	+
			Patients with lumbar disc herniation treated with pharmacotherapy and/or physical therapy (N = 56) (19–21 years, 100% male)[31]	Modified Oswestry Disability Questionnaire	+
			Patients with > 3 month history of chronic low back pain (N = 102) (47.3 ± 14.9 years, 51% male) treated with lumbar fusion (n = 69), decompression (n = 30), or instrument removal (n = 3) [3]	Dallas Pain Questionnaire, work status, Numeric rating scale for back and leg pain	+

TABLE 6. Rating of Overall Strength of Evidence for Each Key Question

Subgroup	Strength of evidence	Conclusions/Comments	Baseline	UPGRADE	DOWNGRADE
Question 1: Do *psychological* subpopulations modify the effect of fusion *versus* conservative management in the treatment of chronic LBP?					
Personality disorder	Insufficient	Patients with a personality disorder may respond more favorably to conservative management and those without a personality disorder more favorably to fusion; however, findings are based on subgroup analyses and not statistically significant.	High	NO	YES (3) Subgroup analyses not stated *a priori* and imprecise estimates
Depression/ neuroticism	Insufficient	Patients with higher depression and neuroticism scores may also respond more favorably to conservative management. These findings need to be confirmed through future clinical research evaluating subgroup effects.			
Question 2: Are there certain screening tests for psychological subgroups that can predict treatment outcome for chronic LBP?					
Descriptive	Not rated	The most commonly cited psychological screening tests in patients with LBP that also have demonstrated predictive validity in the literature include:	NA	NA	NA
		Beck Depression Index (BDI)			
		Fear Avoidance Belief Questionnaire (FABQ)			
		Zung Depression Scale (ZDS)			
		DRAM			
		Spielberger Trait Anxiety Inventory (STAI)			

**An "X" indicates the measure was successfully evaluated for predictive validity.*

Baseline quality: HIGH = majority of article Level I/II. LOW = majority of articles Level III/IV. ***UPGRADE:*** *Large magnitude of effect (1 or 2); Dose response gradient (1) DOWNGRADE: Inconsistency of results (1 or 2); Indirectness of evidence (1 or 2); Imprecision of effect estimates (1 or 2); Subgroup analyses not stated a priori and no test for interaction (2)*

associations between the admission CSQ and the ODI, Symptom Checklist-90 and return to work at discharge from a multidisciplinary pain clinic.[30]

Zung Depression Scale

This 20-item self-report scale measures the four common characteristics of depression: pervasive effect, physiological equivalents, other disturbances, and psychomotor activities. The minimum score is 20 and maximum score is 80. Four categories ranging from "normal" to "severely depressed" are based on specific ranges of the score. Two studies were identified evaluating the predictive validity of this scale. In the first study, 403 volunteers with no history of "serious" low back pain (defined as pain requiring medical attention or absence from work) participated in a functional spinal assessment. At the time of initial assessment and at 6-month intervals thereafter, the volunteers completed the ZDS along with an additional battery of instruments. Scores from the Zung questionnaire were reproducible over 18 months (multiple measurements to assess reliability) and were significant predictors of first time LBP. After accounting for the effects of a history of "nonserious" back pain, psychometric scores predicted less than an additional 3% of reported back pain.

In another study, 102 patients with chronic LBP treated with lumbar fusion (n = 69), decompression (n = 30), or instrument removal (n = 3) were evaluated with a battery of psychological assessment tests including the ZDS 1 to 2 weeks before surgery, and the Dallas Pain Questionnaire, work status, and the numeric rating scale for back and leg pain were assessed 6 months and 1 year after surgery.[3] Regression analyses found a strong predictor of these outcomes to be a combination of the ZDS and MSPQ, known as the DRAM. Patients categorized as "At-Risk" had a twofold higher risk of poor outcome, and patients in either of the "Distressed" categories had a fourfold higher risk of poor outcome.

State-Trait Anxiety Inventory

This 40-item self-report scale provides a measure of *state* anxiety, which is the intensity of anxiety experienced at the time of assessment, and *trait* anxiety, which reflects the general tendency for experiencing anxiety. The two are scored separately and the scores range from 20 to 80 with the higher score representing greater anxiety. A study of 110 outpatients with either acute or chronic low-back pain undergoing conservative treatment completed State-Trait Anxiety Inventory (STAI) along with an additional battery of instruments. Both groups showed elevated state anxiety and those with chronic pain also exhibited mild depression (determined from another measure). Combined scores on depression, anxiety, and negative life change predicted were associated with sensory and affective pain.

Another study consisted of 56 young Korean patients with lumbar disc herniation (LDH) treated with pharmacotherapy and/or physical therapy and 76 controls. All subjects completed the Spielberger's STAI and the BDI.[31] To evaluate pain intensity and functional disability, the Visual Analog Scale and the Modified ODI Questionnaire were used. LDH patients had more depression and anxiety than the controls. The functional disability of the LDH patients was significantly related to the four variables: pain intensity, depression, state anxiety, and trait anxiety. Pain intensity and state anxiety were significantly associated functional disability in the LDH patients.

In another study reviewed earlier for the ZDS, 102 patients with chronic LBP treated with lumbar fusion (n = 69), decompression (n = 30), or instrument removal (n = 3) were evaluated 1 to 2 weeks before surgery. In addition to the ZDS, the STAI was evaluated against the Dallas Pain Questionnaire, work status, and the numeric rating scale for back and leg pain were assessed 6 months and 1 year after surgery.[3] Regression analyses also found the STAI to correlate with outcomes after spine surgery.

Evidence Summary

The overall strength of the evidence evaluating whether specific *psychological* subpopulations modify the effect of fusion *versus* conservative management in the treatment of chronic LBP is "insufficient," that is, evidence either is unavailable or does not permit a conclusion; however, some hypotheses can be generated and considered in clinical decision making and in future research planning (Table 6).

DISCUSSION

The purpose of this systematic review was to determine whether we could identify specific psychological subgroups with chronic LBP that respond more favorably to fusion than to conservative management (or *vice versa*) and to identify the most common psychological screening tests and their predictive validity with respect to LBP outcomes. For the assessment of subgroup treatment effects, we used a methodology that would allow us to evaluate study outcomes on the basis of the heterogeneity of treatment effects. This is best determined by evaluating comparison studies[12–15] that stratify outcomes on patients with different baseline characteristics—what we are calling subgroups. The "textbook findings" for such an analysis would be to find little to no treatment effect comparing two treatments; however, to identify specific baseline characteristics which on the one hand respond more favorably to fusion (*e.g.*, not depressed) and on the other, more favorably to conservative management (*e.g.*, depressed). This can be observed most easily through the use of forest plots. Ultimately, HTE is observed when the treatment effect differences comparing subgroups are statistically significant. This is also known as effect modification and can be tested with a statistical test of interaction.

We identified only one study[32] that compared fusion to conservative management among chronic LBP patients without predominant neurological involvement that stratified outcomes by psychological subgroups. Patients *with a personality disorder* benefited more from conservative management and those *without a personality disorder* benefited more from fusion. However, these heterogeneous responses to treatment were not statistically significant, probably because of the small number of subjects in the personality disorder group; hence, these findings are only hypothesis generating. This study also observed that among patients who improve, the baseline depression and neuroticism scores are significantly higher in the conservatively managed group than the fusion group; however, there is little difference in baseline scores among those who did not improve. These results are not particularly intuitive. It would be far more helpful to see the data presented similar to the personality disorder data so that we could determine if the presence of depression or neuroticism (based on a cutoff value) actually modified the effect of the treatment. These data only allow us to speculate that perhaps patients with higher degrees of depression or neuroticism respond better to conservative management.

With respect to psychological screening tools, we identified six that are most commonly cited in the literature including the BDI, FABQ, CSQ, ZDS, STAI, and the DRAM. Two studies found no association between the CSQ and outcomes for the conservative management of LBP. The remaining four measures (BDI, FABQ, ZDS, STAI, and the DRAM) have been successfully validated against outcomes after treatment for chronic LBP. The FABQ physical activity subscale was associated with the Roland Morris score but not the work subscale. In contrast, the work subscale was associated with the ODI but not the physical subscale in two different studies.

Although there are several case series that report the negative impact of various psychological disorders on CLBP treatment outcomes, there are few level-1 evidence RCT that evaluate this issue between the treatment groups. The study by Hagg *et al*[32] is a good example of the difficulties confronted when attempting to make conclusions on the effects of psychological disorders on the treatment of CLBP. Patients with neuroticism had poor surgical outcomes, while the presence of depression was shown to have little effect. Depression was, however, associated with improved outcomes in the nonsurgical group. Levels of depression, as determined by the ZDS, improved in both groups after treatment, but more so in the surgical group. These findings emphasize the fact that the treatment of pain is complex and can be influenced by many factors beyond the physical findings. In clinical practice, it may be difficult to accurately assess many of the psychological factors that can impact outcomes, and the routine use of a psychological screening questionnaire has been recommended by some.[16] In this analysis, we have identified the most commonly reported, validated, psychological screening questionnaire utilized in the LBP literature. There is not one screening questionnaire that evaluates all potential psychological disorders, but the use of one that has been shown to predict outcomes in the LBP population may be of benefit to the patient and clinician in directing the most efficacious form of treatment.

Strengths of this study include the systematic review approach in identifying comparison studies that reported treatment effects by individual psychological subgroups. This

allowed us to illustrate the significant gaps in the literature with respect to identifying subgroups that may respond more favorably to fusion *versus* conservative management. Despite the case series data suggesting patients with depression or anxiety may not respond as favorably to either treatment, these studies do not provide comparative effectiveness that can assist in treatment decision making since they are being assessed in entirely different study populations and institutions. Such gaps should motivate research to design future trials that also measure subgroup effects. Despite finding that patients with a personality disorder may respond more favorably to conservative management and those without such a disorder more favorably to fusion, one study is not enough to make treatment recommendations, especially since subgroup analyses of secondary data are more appropriately considered hypothesis generating.

Future work in this area should include the analysis of subgroups as part of clinical trials. Subgroup data should be stratified by treatment groups and formal tests of interaction should be performed to confirm the potential of HTE (also known as effect modification). It is our hope that the subgroups we have identified may be further explored with an *a priori* plan to evaluate them in already existing larger databases such as registries. Although any subgroup analysis will have the potential of misinterpretation or spurious findings, nonetheless, such an approach will be very important for future spine research that is aimed at identifying the most important treatment for LBP for each individual patient. This study serves to renew enthusiasm and provide a trajectory for future research efforts aimed at identifying the best treatment for the various subgroups of patients afflicted with chronic back pain. Furthermore, if there are psychological subgroups of patients that respond more favorably to one treatment over another, it is imperative to identify quality baseline screening tools with predictive validity.

When it comes to selecting the appropriate screening tool for pretreatment psychological assessment, one must make several considerations. First, the domain measured (*e.g.*, depression, anxiety, and fear avoidance) and its clinical importance. Findings from this systematic review suggest that depression and patients with a personality disorder are important factors in determining patient treatment and prognosis. Second, a consideration of populations tested in and outcomes tested against is very important. One must also consider patient burden as well. The fewer number of items the better. Both the BDI and ZDS have 21 and 20 items, respectively, while the STAI has 40 items, and the DRAM has 45. The clinician must determine the psychological screening tool that best fits her patient population while considering the burden questionnaires may place on patients and clinicians.

➢ Key Points

- ❑ Very few randomized control trials evaluating the treatment of chronic lower back pain have evaluated the effects of psychological factors on outcomes by treatment intervention.
- ❑ The literature suggests that patients with personality disorders, neuroticism, and depression may have better results with conservative treatment compared with fusion.
- ❑ Six psychological screening tests have been commonly cited in the literature and validated to predict treatment outcomes in the treatment of lower back pain. They are the Beck Depression Inventory, the Fear Avoidance Belief Questionnaire, the Coping Strategies Questionnaire, the Spielberger Trait Anxiety Inventory, the Zung Depression Scale, and the Distress Risk Assessment Method.

Acknowledgment

The authors thank Ms. Nancy Holmes, RN, for her administrative assistance.

References

1. Engel GL. The need for a new medical model: a challenge for biomedicine. *Science* 1977;196:129.
2. Main CJ, Wood PL, Hollis S, et al. The Distress and Risk Assessment Method. A simple patient classification to identify distress and evaluate the risk of poor outcome. *Spine* 1992;17:42.
3. Trief PM, Grant W, Fredrickson B. A prospective study of psychological predictors of lumbar surgery outcome. *Spine* 2000;25:2616.
4. Trief PM, Ploutz-Snyder R, Fredrickon BE. Emotional health predicts pain and function after fusion: A prospective multicenter study. *Spine* 2006;31:823.
5. 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.
6. Coste J, Paolaggi JB, Spira A. Classification of nonspecific low back pain. II. Clinical diversity of organic forms. *Spine (Phila Pa 1976)* 1992;17:1038.
7. Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Phys Ther* 1995;75:470.
8. Hall H, McIntosh G, Boyle C. Effectiveness of a low back pain classification system. *Spine J* 2009;9:648.
9. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q* 2004;82:661.
10. Brookes ST, Whitley E, Peters TJ, et al. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess* 2001;5:1.
11. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 Explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:e1.
12. Brookes ST, Whitely E, Egger M, et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004;57:229.
13. Lagakos SW. The challenge of subgroup analyses—reporting without distorting. *N Engl J Med* 2006;354:1667.
14. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005;365:176.
15. Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med* 2007;357:2189.
16. Daubs MD, Patel AA, Willick SE, et al. Clinical impression *versus* standardized questionnaire: the spinal surgeon's ability to assess psychological distress. *J Bone Joint Surg Am* 2010;92:2878.
17. Grevitt M, Pande K, O'Dowd J, et al. Do first impressions count? A comparison of subjective and psychologic assessment of spinal patients. *Eur Spine J* 1998;7:218.

18. Gatchel RJ, Mayer TG. Psychological evaluation of the spine patient. *J Am Acad Orthop Surg*, 2008;16:107.
19. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am* 2003;85-A:1.
20. Norvell DC, Dettori JR, Fehlings MG, et al. Methodology for the systematic reviews on an evidence based approach for the management of chronic LBP. *Spine* 2011;36: S10–S18.
21. Corporation S. Stata Statistical Software: Release 9.1. College Station, TX: StataCorp LP; 2005
22. Hobby JL, Lutchman LN, Powell JM , et al. The distress and risk assessment method (DRAM). *J Bone Joint Surg Br* 2001; 83:19.
23. Hasenbring M, Marienfeld G, Kuhlendahl D, et al. Risk factors of chronicity in lumbar disc patients. A prospective investigation of biologic, psychologic, and social predictors of therapy outcome. *Spine* 1994;19:2759.
24. Sinikallio S, Aalto T, Airaksinen O, et al. Depressive burden in the preoperative and early recovery phase predicts poorer surgery outcome among lumbar spinal stenosis patients: a one-year prospective follow-up study. *Spine* 2009;34:2573.
25. Al-Obaidi SM, Beattie P, Al-Zoabi B, et al. The relationship of anticipated pain and fear avoidance beliefs to outcome in patients with chronic low back pain who are not receiving workers' compensation. *Spine* 2005;30:1051.
26. George SZ, Fritz JM, Childs JD. Investigation of elevated fear-avoidance beliefs for patients with low back pain: a secondary analysis involving patients enrolled in physical therapy clinical trials. *J Orthop Sports Phys Ther* 2008;38:50.
27. Keeley P, Creed F, Tomenson B, et al. Psychosocial predictors of health-related quality of life and health service utilisation in people with chronic low back pain. *Pain* 2008;135:142.
28. Grotle M, Vollestad NK, Brox JI. Clinical course and impact of fear-avoidance beliefs in low back pain: prospective cohort study of acute and chronic low back pain: II. *Spine* 2006;31:1038.
29. Woby SR, Watson PJ, Roach NK, et al. Coping strategy use: does it predict adjustment to chronic back pain after controlling for catastrophic thinking and self-efficacy for pain control? *J Rehabil Med* 2005;37:100.
30. Dozois DJ, Dobson KS, Wong M, et al. Predictive utility of the CSQ in low back pain: individual *vs*. composite measures. *Pain* 1996;66:171.
31. Kim TS, Pae CU, Hong CK, et al. Interrelationships among pain, disability, and psychological factors in young Korean conscripts with lumbar disc herniation. *Mil Med* 2006;171:1113.
32. Hagg O, Fritzell P, Ekselius L, et al. Predictors of outcome in fusion surgery for chronic low back pain. A report from the Swedish Lumbar Spine Study. *Eur Spine J* 2003;12:22.
33. Ackerman MD, Stevens MJ. Acute and chronic pain: pain dimensions and psychological status. *J Clin Psychol* 1989;45:223.
34. Mannion AF, Dolan P, Adams MA. Psychological questionnaires: do "abnormal" scores precede or follow first-time low back pain? *Spine* 1996;21:2603.