Published in final edited form as:

Neuromodulation. 2008 October 8; 11(4): 237–248. doi:10.1111/j.1525-1403.2008.00171.x.

A DEMONSTRATION OF A PRESURGICAL BEHAVIORAL MEDICINE EVALUATION FOR CATEGORIZING PATIENTS FOR IMPLANTABLE THERAPIES: A PRELIMINARY STUDY

Kimberly Gardner Schocket, Ph.D., Robert J. Gatchel, Ph.D., Anna Wright Stowell, Ph.D., Martin Deschner, Ph.D., Richard Robinson, Ph.D., Leland Lou, M.D., Tony Whitworth, M.D., and Dana Bernstein, M.S., Ph.D.

The Eugene McDermott Center for Pain Management, University of Texas Southwestern Medical School at Dallas, TX

Abstract

Objectives—The aim of the current study was to evaluate the potential efficacy of a presurgical behavioral medicine evaluation (PBME) screening algorithm with patients undergoing evaluation for implantable pain management devices.

Methods—Sixty patients were evaluated for prognostic recommendations regarding outcomes from surgery for spinal cord stimulators and intrathecal pumps. Diagnostic interviews, review of medical charts, and psychosocial and functional measures were used in the initial evaluation.

Results—Patients were classified into one of four prognostic groups, from low to increasing risks: Green, Yellow-I, Yellow-II, and Red. The Green group showed the most positive biopsychosocial profile, while the Red groups showed the worst profiles.

Conclusions—This preliminary study suggests that the PBME algorithm may be an effective method for categorizing patients into prognostic groups. Psychological and adverse clinical features appear to have the most power in the classification of such patients.

Keywords

implantable devices; presurgical screening algorithm; psychosocial risk factors; adverse clinical features

INTRODUCTION

Pain is one of the most costly epidemics facing society today causing untold suffering, diminished quality of life, and enormous economic ramifications for millions of Americans. It is estimated that one out of every 14 people seek medical care for back or neck pain, resulting in over 14 million pain related health-care visits annually, with a total cost exceeding \$70 billion per year. Thus, the importance of finding efficacious modalities for the treatment of chronic pain is unrivaled.

Many chronic pain patients are refractory to conservative medical therapy. After failing traditional therapies, such as opioids and other minimally invasive procedures, these patients are often selected for invasive pain management techniques. Spinal cord stimulators (SCS) and intrathecal (IT) opioid delivery systems are two semi-permanent options for these

patients. However, they require a large economic input, as well as a rather invasive placement procedure¹. Cost-effective analyses show SCS and IT pumps to be more costly at onset¹ but, after 2.5 years, they become less costly than conventional pain therapies^{1, 2}. With success rates greater than 50%^{3, 4}, these are very powerful tools for the physician; however, due to the high costs, it would be extremely beneficial to accurately predict those patients who will have good outcomes.

There is a copious amount of literature examining the predictive value of biopsychosocial risk factors in surgical outcomes^{5–9}. Yet, it has only been in the last decade that studies have begun to look at the relationship between behavioral medicine risk factors and treatment outcomes for implantable modalities^{10–12}. These reviews show psychosocial factors have a large impact on treatment outcomes. The most comprehensive method to date for the examination of psychosocial and medical risk factors was developed by Block and colleagues^{5, 13, 14}. This presurgical psychological screening protocol has been shown to be effective in predicting general spine surgery outcomes¹⁴. The current study evaluated whether the application of Block's presurgical screening method would be useful in categorizing patients' potential suitability as candidates for a subset of implantable modalities, specifically SCS and IT opioid systems.

The present study was designed to be the first preliminary evaluation of whether the application of the algorithm, previously validated only for spine surgery candidates, could potentially generalize to a subset of patients undergoing presurgical behavioral medicine screening for implantable devices to manage chronic pain, specifically SCS and IT pumps. The goal was to demonstrate the evaluation process, and take the important step of simply assessing the psychosocial differences among the resultant prognostic pre-screening categories.

Of course, one might argue that, because neuromodulation is quite different from spine surgical intervention, this algorithm would not be applicable to the former. However, for any type of presurgical screening that involves stress and uncertainty, such as the current one, the global psychological resilience of the patient (including the psychological strengths and weaknesses in dealing/coping with stress) is the key entity that is evaluated. That is to say, the global capacity of the patient to respond adaptively, whatever the specific challenges of any particular surgery, is the key ingredient. With this in mind, it was expected that Block's algorithm would be clinically applicable to neuromodulation procedures. We attempted to delineate differences found among prognostic groups with regard to psychosocial, functional, and medical risk factors. Prospectively, we hypothesized that demographic variables, namely disability compensation, would be significant in determining group recommendation assignments. It was also hypothesized that the Green group (least number of risk factors) would show the best biopsychosocial profile at initial evaluation when compared to all other groups. Additionally, we hypothesized that psychological test data would have the most power with regard to prognostic assignments. Lastly, we hypothesized that the Green group would show the best biopsychosocial profile at six-month post initial evaluation, and the Red group (highest number of risk factors) would show the worst. These hypotheses were assessed using various physical/functional and psychosocial outcome measures.

METHOD

Subjects

Patients were referred to an interdisciplinary pain center for Presurgical Behavioral Medicine Evaluations (PBME) by physicians seeking recommendations regarding surgery for implantable devices to help control chronic pain. The specific elements of the PBME are

clearly delineated by Block and colleagues⁵. Sixty consecutively selected patients, who completed a PBME, were included in the study. The total study sample was composed of 35 (58.3%) women and 25 (41.7%) men; mean age was 56 years, with a range of 30 to 90.

Procedure

Because the PBME procedure and any resultant surgery was part of patients' standard care at The Eugene McDermott Center for Pain Management (The Center), no Institutional Review Board involvement was required. A trialing mechanism and implant criteria were followed for both SCS and IT patients. For SCS patients, a trial consisted of implanting one or two percutaneous leads for a test period of 3 – 5 days. Those patients who reported at least a 50% reduction in pain were considered for SCS implant. For IT patients, the trial consisted of a single bolus injection of 1 mg morphine intrathecally. Patients who reported at least a 50% reduction in their pain levels were then considered for IT implant. Prior to their scheduled evaluation with the clinical psychologist, patients received a packet of information about the PBME. This packet included an explanation of the upcoming evaluation, a consent form for psychological assessment and testing, and questionnaires evaluating pain levels, medication usage, impact of pain on physical and emotional abilities, and overall impact of pain on lifestyle. At the initial evaluation, the informed consent was explained and obtained from each patient. Following this, each patient participated in a diagnostic clinical interview (approximately 90 minutes) and completed various instruments assessing psychosocial and physical functioning (approximately 3 hours).

Tests Administered—The following tests were administered to all subjects: a general pain questionnaire, which asked questions concerning duration of pain, past interventions, etc.; the Visual Analogue Scale (VAS), which is designed to rate the patient's degree of pain on a scale from 0 (no pain) to 10 (worst possible pain)¹⁵; the Million Visual Analog Scale¹⁶ is comprised of 15 self-report measures, with subjects indicating their response on a 10 centimeter line, with a total score ranging from 0 (no pain/disability) to 150 (maximum pain/disability); the Oswestry Disability Questionnaire (OSW¹⁷) is comprised of 10 questions, with each item scored on a 0 – 5 point scale, with a potential score of 0 (no disability) to 50 (severe disability); the Pain Medication Questionnaire (PMQ¹⁸) is a screening tool developed as an instrument to assess risk of opioid medication misuse. It consists of 20 self-report items, with each item scored on a standard 5-point Likert scale, a minimum of 0 and a maximum score of 104 can be obtained; the Beck Depression Inventory (BDI¹⁹) is a 21-item self-report inventory that assesses the intensity of depressive symptomatology. Each item is scored from 0 to 3, with a potential range from 0 to 63; the Millon Behavioral Medicine Diagnostic (MBMD²⁰) is a 165-item self-report questionnaire, consisting of 29 clinical scales, 3 response pattern scales, and 1 validity indicator, and 6 negative health habit indicators; the Minnesota Multiphasic Personality Inventory-Second Edition (MMPI-2²¹) is a 567-item, self-report measure of personality functioning and psychiatric symptoms that yields 10 clinical scales, numerous supplementary scales, and several validity scales; the Medical Outcomes Short Form Health Survey (SF-36) developed by Ware et al.²² is a 36-item questionnaire that assess health-related quality of life (both physical and mental) from the point of view of the health-care recipient. A “normal score” is 50, with a standard deviation of 10. Lower scores reflect poorer quality of life; the Coping Strategies Questionnaire (CSQ²³) is a 42-item self-report questionnaire, with each item scored from 0 – 6. The Catastrophizing scale consists of 6 items; the Hamilton Psychiatric Rating Scale for Depression (HAM-D²⁴) consists of 17 items, rated on a 3-to-5 point scale, and ranges from 0 to 68; and the Treatment Helpfulness Questionnaire (THQ²⁵) is an 11-item patient-rated inventory of helpfulness of various aspects of treatment. Patients rate the particular modality on a 16-point scale, in which 0 = very harmful, 4 = harmful, 8 = neutral, 12 = helpful, and

16 = very helpful. Additionally, each patient's medical chart was reviewed. The referring physician also completed a medication assessment to report on patients' medication regimen.

Categorization—The revised algorithm included both psychosocial and medical risk factors⁵. The psychosocial risk factors included: job dissatisfaction, worker's compensation status, pending litigation, spousal solicitousness or lack of support, abuse and/or abandonment, substance abuse, history of psychological disturbance, pain sensitivity, chronic and/or reactive depression, anger, anxiety, and a pathological depression profile (see reference ⁶ for details). The medical risk factors were: duration of pain, type of surgery, presence of non-organic signs, abnormal pain drawing, previous surgeries, prior medical problems, smoking, and obesity (again, see reference ⁶ for details). The results from the diagnostic interview, medical chart review, and psychosocial/functional testing were integrated using Block's algorithm, which lead to each patient's prognostic classification. Again, we used the Block algorithm to classify patients into one of three categories: Good (our Green group), Fair (our Yellow group), or Poor (our Red group). In addition, we subsequently made additional subcategories (Yellow I and Yellow II, as well as Red I and Red II) in order to evaluate whether they would provide more sensitivity. These results were given to the referring physician, delineating the basic problem areas and the recommendations for surgery.

Recommendations—The psychologists' recommendations fell into one of five categories: 1) proceed with surgery (Green); 2) surgery with post-operative psychological sessions (Yellow-I); 3) pre-operative psychological sessions prior to surgery (Yellow-II); 4) non-invasive therapy recommended (Red-I); 5) no treatment of any kind (Red-II). The referring physician then followed up with the patient to discuss a surgery and/or treatment plan. Patients were also given the option of making an additional appointment with the psychologist to discuss the results from the PBME directly.

Those patients falling into the recommendation groups, where psychological treatment (pre-operative or post-operative) was recommended, were given the option of proceeding with that treatment. Pre-operative treatment typically consisted of three - four behavioral medicine sessions with a psychologist to help prepare the patients to manage the psychosocial factors that can influence recovery after surgery. Upon completion of these sessions, the patients were given revised surgery recommendations. Post-operative treatment (between 1 and 10 sessions) focused on compliance and motivation to help the patient cope and adjust to issues that arise after implantation. Additionally, an attempt was made to contact patients for a follow-up telephone interview six-months post initial evaluation. The follow-up interviews included testing measures (BDI, SF-36, OSW) and brief questions assessing pain level, surgery status, vocational status, healthcare utilization, and changes in demographic variables. Referring physicians also completed a medication assessment at six-month post-initial evaluation regarding changes in medications and/or dosages.

RESULTS

For statistical analyses, the five prognostic categories were collapsed into four groups due to the small sample size of the Red-I and Red-II groups. The four categories include: Green, Yellow-I, Yellow-II, and Red (combined Red-I and Red-II). All statistical analyses were computed using SPSS®, Version 13.

Demographic Characteristics

The demographic characteristics of the four prognostic groups are presented in Tables 1A and 1B. Significant differences were found for gender ($p = .03$) and disability payment

status ($p = .05$). Pairwise comparisons revealed that male patients were 18.7 times more likely than females to be classified in the Green vs. Red group, $\chi^2(1) = 6.74$, $p = .009$, OR = 18.7, 95% C.I.: 1.56–222.93. Regarding disability payment status, pairwise comparisons revealed that the Red group (55.6%) contained significantly greater proportions of patients receiving disability when compared to the Green group (0%). Additionally, a significant linear trend was observed among the groups with regard to disability payment status ($p < .01$), with more disability payments associated with increasing risk status. No significant differences were found among the four prognostic groups on the variables of age, race, marital status, pending litigation, or pain duration.

Physical/Functional Measures at Initial Evaluation

The physical/functional measures were analyzed using ANOVAs, with post-hoc comparisons to identify specific differences across the groups. No significant differences were found among groups on the measures of perceived pain status (VAS), perceived pain and disability (MVAS), and physical-health related impairment (SF-36/PCS). However, significant differences were seen among the four groups on the OSW, a measure of patient-perceived pain related physical limitation, $F(3, 48) = 3.63$, $p = .019$, where the higher the score, the greater the limitation. Tukey HSD test/corrections showed the Red group endorsing significantly more limitations of activities of daily living (ADL) and disability than the Green group, followed by the Yellow-II group also endorsing limited ADLs and disability (Figure 1).

Psychosocial Measures at Initial Evaluation: Coping Measures

Coping Strategy Questionnaire (CSQ)—The CSQ examines six different types of coping strategies: 1) pain diverting attention; 2) reinterpreting pain sensations; 3) coping self-statements; 4) ignoring pain sensations; 5) praying or hoping; and 6) catastrophizing. Catastrophizing can be broadly defined as “an exaggerated negative mental set brought to bear during actual or anticipated pain experience.” Individuals who tend to catastrophize experience higher levels of psychological distress, poorer physical functioning and increased disability, and greater levels of pain intensity. The CSQ was analyzed using ANOVAs, with post-hoc comparisons to specify the differences among the groups. No significant differences were found on the CSQ total score, a measure of overall level of coping strategy employed to manage pain. However, a significant difference was observed on the Catastrophizing scale of the CSQ, $F(3, 48) = 7.06$, $p = .001$. Pairwise comparisons revealed that the Red group scored significantly higher on the Catastrophizing scale than the three other groups (Table 2).

Mental Component Scale (SF-36/MCS)—Significant differences were found on the MCS, a measure of mental functioning, $F(3, 16) = 4.72$, $p = .015$. Particularly, the Red group's score, the lowest score on the MCS among the four groups, was significantly lower than the Yellow-I group's score (Table 2).

Psychosocial Measures at Initial Evaluation: Mood and Personality Measures

Hamilton Psychiatric Rating Scale for Depression (HAM-D)—At initial evaluation, significant differences were found among groups on the HAM-D, $F(3, 56) = 17.14$, $p < .001$. Post-hoc analyses revealed that the Green and Yellow-I groups scored significantly lower (scores in the none-minimal range) than the Yellow-II and Red groups. There were also significant differences seen between the Yellow-II and Red groups, with the Red group scoring significantly higher on the HAM-D, reflecting depressive symptomatology falling in the moderate-severe range (Table 2).

Beck Depression Inventory (BDI)—Significant differences were also found on the BDI, a measure of depressive symptomatology, $F(3, 56) = 12.53$, $p < .001$. Once more, Tukey HSD test/corrections revealed that the Red group scored significantly higher on the BDI than all other groups. The Green group scored the lowest, endorsing the fewest symptoms of depression, followed by the Yellow-I and Yellow-II groups, respectively (Table 2).

Psychosocial Measures at Initial Evaluation: Personality Measures

Minnesota Multiphasic Personality Inventory-Second Edition (MMPI-2)—The MMPI-2 clinical scales were utilized to examine the relationship between psychological functioning and prognostic determination. A MANOVA was first performed to determine if significant differences existed among the overall mean T-scores of the four prognostic groups on the MMPI-2 clinical scales. Significant differences were observed among the four groups and the MMPI-2 clinical scales, Hotelling's Trace = 1.44, $F(39, 122) = 1.51$, $p = .048$, justifying conducting ANOVAs for each individual MMPI-2 scale. Differences were seen among prognostic groups for several clinical scales including the F Scale ($p < .001$), the K (Correction) Scale ($p = .033$), Scale 1 Hypochondriasis ($p = .004$), Scale 2 Depression ($p = .016$), Scale 3 Hysteria ($p = .030$), Scale 4 Psychopathic Deviate ($p = .006$), Scale 6 Paranoia ($p < .001$), Scale 7 Psychasthenia ($p = .008$), Scale 8 Schizophrenia ($p < .001$), and Scale 9 Hypomania ($p = .022$). Planned contrast analyses indicated that the Green and Yellow-I groups together had significantly lower mean scores than the Yellow-II and Red groups together. They also showed that the Green group had a significantly higher mean score on the K (Correction) Scale than the Red group, and a significantly lower mean score on Scale 0 Social Introversion than the Red group (Figure 2).

Million Behavioral Medicine Diagnostic (MBMD)—The MBMD is designed to assess psychological factors that can affect the course of medical treatment and recovery. A MANOVA examined the relationship among the 29 MBMD clinical scales and the 4 prognostic groups, and found a significant scale effect, Hotelling's Trace = 6.27, $F(29, 87) = 1.78$, $p < .01$. These findings suggest that univariate analyses of variance may be safely conducted for each individual scale, without undue inflation of Type I error rates. Results of the subsequent ANOVAs revealed significant differences among the prognosis groups on the following scales: Anxiety Tension ($p = .006$), Depression ($p < .001$), Cognitive Dysfunction ($p = .004$), Emotional Lability ($p = .005$), Inhibited ($p < .001$), Dejected ($p < .001$), Cooperative ($p = .003$), Confident ($p = .002$), Nonconforming ($p = .041$), Oppositional ($p < .001$), Denigrated ($p < .001$), Illness Apprehension ($p = .016$), Functional Deficits ($p = .028$), Pain Sensitivity ($p = .010$), Social Isolation ($p = .001$), Interventional Fragility ($p = .006$), Information Discomfort ($p = .007$), Utilization Excess ($p < .001$), Adjustment Difficulties ($p = .002$), and Psych Referral ($p = .001$). Planned contrasts and post-hoc analyses revealed that the differences were largest between the Green and Red prognostic groups for most of the significant scales, with the Green group having significantly lower mean scores when compared to the Red group.

Medication Use at Initial Evaluation

Physician Medication Assessment—The Physician Medication Assessment was analyzed using complex chi-squares to determine if differences existed in medication use among the groups. Medication was broken down into three classifications, including narcotic use, non-narcotic use, and no medication use. No differences in medication use were seen at initial evaluation among the groups.

Healthcare Utilization at Initial Evaluation

The number of healthcare visits and emergency room visits in the past year were analyzed relative to prognostic group. A one-way ANOVA revealed significant differences among groups with regard to the number of healthcare visits in the 6 months prior to initial evaluation, $F(3, 40) = 4.21$, $p = .011$, but showed no differences in number of ER visits in the 6 months prior to the initial evaluation. The Red group ($M = 24.5$, $SD = 24.0$) had 6 times more healthcare visits in the 6 months prior to initial evaluation than the Green group ($M = 4.13$, $SD = 3.0$).

Vocational Status at Initial Evaluation

Pearson Chi-Square analyses were performed to examine vocational status among the four prognostic groups at initial evaluation. The patients were classified into one of the following categories: currently working; not working due to original injury; and not working due to reasons unrelated to original injury. No significant differences were found among prognostic groups with regard to vocational status. However, further examination, using the Mantel-Haenszel statistic, showed a linear trend, $X^2(1) = 5.74$, $p = .017$. The Green group had the highest percentage of patients currently working (40%), with the number declining with prognostic level: Yellow-I (19%), Yellow-II (10.5%), Red (0%). In addition, the Red group had the highest percentage of patients not working due to the original injury (66.7%) versus the other groups: Yellow-II (57.9%), yellow I (47.6%), Green (20%).

Specific Algorithm Component Scores Relative to Prognostic Groups

The current study determined which prognostic group patients would fall into based on overall algorithm risk scores. The total risk score for each component of the overall algorithm—interview data, psychological test data, and medical factors -- were then evaluated using predetermined *a priori* weights for each component. One-way ANOVAs and post-hoc comparisons were employed to examine differences among prognostic groups on these overall component risk scores.

During the clinical interview, patients reported on factors such as level of job satisfaction, workers' compensation status, pending litigation related to their pain, history of abuse or abandonment, substance abuse, and psychological history. They also reported on the amount of spousal support and/or solicitude they received. A one-way ANOVA revealed that there were significant differences among groups at initial evaluation, $F(3, 56) = 6.42$, $p = .001$. Tukey HSD pairwise comparisons indicated that the Green and Yellow-I groups had significantly lower interview risk scores when compared to the Red group ($p < .01$). Differences were also found between the Green and Yellow-II groups ($p < .05$).

The psychological measures used at initial evaluation included the BDI, MMPI-2, HAMD, CSQ, and MBMD. These tests yielded information about the patients' level of pain sensitivity, depression, anxiety, and catastrophizing. A one-way ANOVA showed significant differences among groups on the overall psychological risk score, $F(3, 56) = 6.79$, $p = .001$. Tukey HSD pairwise comparisons revealed significant differences existed between the Green group when compared to both the Yellow-II and Red groups ($p < .01$). Differences were also found between the Green and Yellow-I group ($p < .05$). Thus, the Green group showed a significantly lower overall risk scores on psychological measures than the other three groups.

Similar to the interview and psychological risk scores, one-way ANOVAs showed significant differences existed among groups relative to medical risk factors, $F(3, 56) = 3.12$, $p = .033$. These factors include duration of pain, number and type of prior spine surgeries, nonorganic physical signs, abnormal pain drawings, smoking, and obesity. Tukey

HSD test/corrections showed the differences to exist between the Green and Red groups, $p = .031$, with Green group scoring significantly lower overall medical risk scores when compared to the Red group.

The overall adverse clinical features risk score, according to Block's algorithm, was based on the presence or absence of any adverse clinical features. Adverse clinical features contributing to the algorithm include inconsistency, medication seeking, staff splitting, compliance issues, threatening, resignation, deception, and personality disorders. Pearson Chi-Square analyses showed significant differences existed among groups, relative to the total presence/absence of these adverse clinical features, $X^2(1) = 22.76$, $p < .001$ (Table 3). Planned contrasts analyses, conducted to further examine these differences, showed the Green group to be 31.5 times more likely to have no adverse clinical features when compared to the Red group, $X^2(1) = 8.93$, $p = .003$, OR = 31.5, 95% C.I.: 2.35–422.30. The Yellow-I group proved to be 70 times more likely to be absent of adverse clinical features when compared to the Red group, $X^2(1) = 17.18$, $p < .001$, OR = 70.00, 95% C.I.: 5.47–896.59. Similarly, the Yellow-II group proved to be 19.8 times more likely to be found without adverse clinical features when compared to the Red group, $X^2(1) = 10.83$, $p = .001$, OR = 19.83, 95% C.I.: 2.70–145.67.

In order to more fully evaluate the impact of adverse clinical features, we built upon Block's original algorithm and scored each adverse clinical feature with *a priori* weights. Patients were given a score between 0 and 2 (0 = absence of adverse clinical features, 1 = presence of adverse clinical features with mild significance, and 2 = presence of adverse clinical features with moderate to severe significance) for each of the 8 adverse clinical features, with a total possible score of 16. The cumulative scores were then analyzed using an ANOVA. These analyses yielded significant differences among the four prognostic groups ($p < .001$). Further, Tukey HSD test corrections showed the Red group to have a significantly higher mean score than all other groups ($p < .001$) (Figure 3).

Factors Associated with Prognostic Groups

Binary logistic regression models were developed for each prognostic group to elucidate the factors that are most associated with group membership. Factors were entered into the regression equation if statistical differences were found in previous analyses. Four factors were found to be associated with Green group membership, with 90.4% accuracy (95.3% sensitivity and 66.7% specificity). They include the BDI total score, the CSQ Catastrophizing scale, the OSW total score, and the interview risk score. The SF-36/MCS and HAM-D total score were the two factors found to be associated with Yellow-I status in the regression model, with 90.0% accuracy (100% sensitivity and 75.0% specificity). The Yellow-II group also was found to have two significant factors, the SF-36/MCS and the BDI total score, with 80% accuracy (93.3% specificity and 40.0% sensitivity). The Red group model was found to be most associated with accuracy at 95.0% (96.1% sensitivity and 88.9% specificity). The factors found to be associated with the Red group membership included the BDI total score and the presence of adverse clinical features.

Follow-Up Data

Although this study did not provide the opportunity for long-term follow-up in most patients, data were collected from a small cohort of patients who were able to complete the six-month follow-up evaluation. The current study hypothesized that the Green group would have a significantly better biopsychosocial profile at six-month follow-up when compared to the other prognostic groups. Overall, the Green group was expected to have significantly lower scores on both physical/functional measures (with the exception of the PCS, where a higher score indicates more positive physical functioning) and psychosocial measures (with

the exception of the MCS, where a higher score indicates more positive mental functioning) at the six-month point. Additionally, it was anticipated that differences would be found among the prognosis groups on similar physical/functional and psychosocial measures. Healthcare utilization, vocational status, and medication use, relative to prognosis group, were also compared. At the time of this project, 34 patients reached the 6-month follow-up and, of those, 31 completed the follow-up evaluation. Relocation, noncompliance, and intervening medical conditions were the reasons cited for those patients unable to complete the six-month follow-up. Of the 31 completed follow-up evaluations, 20 patients underwent surgery for implantable devices; 25% received SCS and 37.5% received IT morphine pumps, while the rest of the patients had surgery unrelated to their pain (6.3%) or did not have surgery (31.3%).

Physical/Functional and Psychosocial Measures—In order to further evaluate group differences on these measures, ANCOVAs and repeated measures ANOVAs were used to further assess the improvements from initial evaluation to six-month follow-up, as well as to determine whether the four groups differed from one another biopsychosocially. Given the significant differences found at initial evaluation on the BDI, OSW, and MCS, one-way ANCOVAs, with pretreatment scores as covariates, were conducted to determine if there were differences among the four prognostic groups on these physical/functional and psychosocial measures. Analyses yielded no significant differences in improvement on these measures among the prognostic groups. It is important to note that the small sample sizes at six-month follow-up most likely affected the statistical power of these analyses to detect significant differences.

Change in Medication Use—Change in medication use within each prognostic group was analyzed using the Wilcoxon sign ranks test. Medication use was coded into three groups at both initial evaluation and six-month follow-up, including narcotic use, non-narcotic use, and no medication. Significant differences were found in the Yellow-I ($z = -2.12, p = .034$) and Yellow-II ($z = -2.25, p = .024$) groups. Both Yellow groups showed improvements in the overall medication use, as evidenced by reduced narcotic use, change from narcotic to non-narcotic medication, and/or change to no medication. The small sample size in the Green and Red groups likely affected power, and these groups did not show significant differences in medication usage (Figure 4). Complex chi-square analyses were also conducted to determine if differences in medication use existed among the groups at six-month follow-up. Differences were not statistically significant; however, a trend was observed among the groups with regard to no medication use. The Green group showed 40% of patients to be taking no medication at six-month follow-up, and this percentage decreased as prognosis worsened (Yellow-I 27.3%, Yellow-II 25%, and Red 0%).

DISCUSSION

Previous analyses of the Block et al.⁵ presurgical screening algorithm proved to be extremely effective in predicting outcome for patients undergoing spine surgery. The current study was designed to take an important first step in evaluating whether this screening algorithm has potential utility for a subset of patients being assessed for implantable devices for pain management, namely spinal cord stimulators and IT opioid delivery systems. If comparable psychosocial differences were found among the four risk categories in the present patient cohort as previously displayed by the spine patient cohort, then the potential utility of its use would be supported (and, thus, warrant further evaluation in future research). Analyses were conducted at initial evaluation to determine if significant differences in demographic variables existed among the four prognostic groups. The first hypothesis of the current study predicted that particular demographic variables, namely disability payment status, would be statistically significant among the group assignments.

Among the variables of gender, race, marital status, disability payment status, and litigation status, only gender and disability payment status demonstrated significant differences in distribution among the Green, Yellow-I, Yellow-II, and Red groups. Male patients were 18.7 times more likely than females to be classified in the Green prognosis group, and 5.4 times more likely to fall in the Green group over females when comparing Green and Yellow-II groups. For disability payment status, the Red group contained a significantly greater proportion of patients receiving disability monies than all other groups. A significant linear trend was observed, showing a correlation between prognosis group and disability payments, where the number of patients receiving disability payments increased as prognosis worsened from Green (0%) to Red (55.6%). This finding supports the idea that patients receiving disability money tend to have poorer surgical results and overall outcomes^{26–28}.

The second hypothesis of the current study predicted that patients in the Green group would show better biopsychosocial functioning than all other prognosis groups at initial evaluation. Of course, this was not surprising because Block's algorithm requires the classification be based on variables such as depression, anxiety, worker's compensation status, etc. Nevertheless, physical and functional measures showed the Red and Yellow-II groups endorsed significantly more limitations on ADLs and disability than the Green group, which endorsed the least amount of limitations. Therefore, as expected, perceived physical functioning of patients in the Green group was higher than all other groups. With regard to vocational status, a linear trend analysis showed the Green group had the greatest number of patients currently working, followed by the Yellow-I and Yellow-II groups. The Red group did not contain any patients who were working at initial evaluation. Employment status is a good indication of functionality for patients. Examination of healthcare utilization in the six months prior to the initial evaluation yielded significant results, with the Red group demonstrating six times more healthcare visits than the Green group.

Psychological and social distress is cited as having significant impact on surgical results¹⁴. In the current study, patients in the Green and Yellow-I groups showed significantly lower mean scores on multiple psychological measures when compared to Yellow-II and Red groups. The Yellow-II and Red groups showed elevations on MMPI-2 scales that have been shown to correlate with diminished surgical results, including the F Scale, Scale 1 Hypochondriasis, Scale 2 Depression, Scale 3 Hysteria, Scale 4 Psychopathic Deviate, Scale 6 Paranoia, Scale 7 Psychasthenia, and Scale 8 Schizophrenia. Specifically, Scale 1 and Scale 3 elevations have been correlated with poorer surgical outcomes. In this study, the Red group, followed by the Yellow-II group, showed the highest elevations on these scales.

Patients exhibiting symptomatology of depression have consistently been shown to have poor results after surgery. Patients in the Green group showed significantly lower mean scores on Scale 2 of the MMPI-2 than those in the Red group, indicating lower levels of depressive symptomatology. The BDI ($p < .001$) and the HAM-D ($p < .001$) were also consistent, showing the Green group as having significantly lower levels of depressive symptoms (significant linear trends on both BDI and HAM-D).

Scale 4 of the MMPI-2 is the best approximate of a patient's anger, which has been shown to have negative effects on surgical outcome and recovery. The Green group showed the lowest scores on Scale 4, exhibiting the least amount of anger among the groups and indicating they would have the greatest likelihood for success of surgery. Anxiety has also been linked to poor surgical outcomes. As expected, the Green group demonstrated the lowest levels of anxiety as measured by the MMPI-2. Research has extensively linked spousal and social support to better overall outcomes and recovery. The Green group also showed a significantly lower mean score on Scale 0 Social Introversion when compared to the Red group. These differences were expected, as the MMPI-2, BDI and HAM-D was

used as part of the screening algorithm to determine in which prognostic group patients were placed. Parallel to the psychosocial measures mentioned above, several of the MBMD clinical scales also showed the Green group to have the lowest mean scores, signifying minimal symptomatology, with the Red group having significantly higher mean scores.

The way in which individuals deal with stress has also been correlated to surgical outcomes. The Red group was shown to have the highest scores on the CSQ Catastrophizing scale, measuring the overall level of the catastrophizing coping strategy employed by an individual to help manage pain. A linear trend was observed indicating that, as prognosis worsened, catastrophizing increased. Thus, the Green group showed the best overall biopsychosocial functioning, and the Red group showed the worst biopsychosocial functioning when compared to the other prognostic groups. Additionally, the Red group indicated having the greatest amount of mental impairment when compared to the other groups, as measured by the SF-36/MCS.

Analysis of the four overall algorithm risk scores showed the interview risk score, psychological test risk score, and adverse clinical features risk score as significantly different across groups. The differences seen in the medical risk factor score were the least significant. These findings are similar to those of Block and colleagues in their study of the algorithm¹⁴, where they found medical risk factors to contribute the least to the overall predictive value of the presurgical algorithm. An additional analysis using a revised scoring system for the adverse clinical features, where each feature was scored based on assigned *a priori* weights, yielded significant differences among groups. The Red group showed a significantly greater mean score than all other groups. This scoring system was not used in the original Block algorithm; the original algorithm considered either the presence of adverse clinical features (score of 1) or the absence of adverse clinical features (score of 0). This revised method of scoring adverse clinical features allows for a more comprehensive examination of each group with regard to adverse clinical features, which proved extremely significant for prognostic purposes, especially with poor surgery candidates. It should be noted that, even though these data are somewhat preliminary in nature and therefore do not fully establish the reliability of a four category solution to a potential prognostic model, they represent a vitally important first step in more fully developing a truly reliable and valid model. This awaits further investigation.

The final hypothesis of this study predicted that psychological test data would have the most predictive value in determining which patients would fall into the four prognostic groups. Regression analyses for each prognostic group revealed that psychological test data were indeed the most predictive. The regression model that most predicted the Green group included the BDI, the CSQ Catastrophizing scale, the OSW, and the overall interview risk score. Such findings suggest that these variables may be used to develop a shorter version of the PBME. This possibly awaits further investigation. The Yellow-I group model indicated that the MCS and the HAM-D scores were most predictive of membership. The MCS and BDI were found to be most predictive in the Yellow-II group model. Lastly, the Red group model, found to have the highest accuracy rate (95%), revealed the BDI and the presence of adverse clinical features to be most predictive of membership in the Red group. As seen by these analyses, the BDI was used in three of the four models, the MCS used in two models, the HAM-D and CSQ Catastrophizing scale were also present in models; thus, psychological test data impacted the results of the group prognosis more than all other factors. The level of depression and overall mental impairment endorsed by patients proved to be significantly associated with overall prognosis. Lower levels of depression were associated with positive outcomes, with higher levels being more indicative of worse outcomes.

Adverse clinical features were surprisingly found to be one of the two factors in the Red group's regression model. These features have not been empirically researched; however, they appear to be important, in that they were significantly associated with poor prognosis. The revised scoring methodology used in this study for additional analyses with the adverse clinical features was able to better depict differences among the groups, especially the poor surgery candidates. Adverse clinical features can be described as aspects of each patient's case that can be identified as having potential to negatively affect outcomes⁵. They include things such as inconsistency, medication seeking, staff splitting, and deception. This is an important finding in that these features, added to those of Block and colleagues in 2003, may be more important to prediction methodology than originally thought. Revising the methodology used to account for the adverse clinical features is beneficial for more comprehensive identification of the differences among groups.

Of course, as with most clinical research endeavors, this present study is not without some limitations. One limitation was the relatively small sample size for the initial and 6-month follow-up analyses. The small number of patients in each group most likely hindered the power of the 6-month follow-up analyses, making it very difficult to find statistical differences among the groups. Several of the outcome analyses showed improvements within and between groups; however, the sample size in most cases was not substantial enough to create a statistically significant effect. The distribution of patients among groups was also a limitation, as most patients were clustered in the Yellow-I and Yellow-II groups. This is natural in that most patients fall into the fair (middle) prognostic categories rather than the good and poor. It should also be noted that a large number of statistical analyses were conducted in the present study, thus possibly increasing the likelihood of Type I errors. However, as Aicken and Gensler²⁹ point out, there is still substantial debate in the biostatistical and epidemiological literature concerning whether adjustment for multiple tests (such as the Bonferroni procedure) is warranted, especially when preliminary or exploratory analyses are being conducted. Because we have performed many preliminary analyses of data that have not been evaluated in the past, we have avoided being too conservative in our analyses in the present study. Finally, a potential limitation of the PBME is the fact that it does take additional time and requires the use of a trained clinician to administer.

Another question that can be raised concerning this study was the fact that, unlike spinal surgery for which this algorithm was specifically developed, it may not be totally generalizable to an implant therapy population where there is a pre-implant trial administered to assess potential efficacy and needed calibration adjustment before the final implantation. Future studies will need to conduct more systematic follow-ups of patients to evaluate the degree to which the PBME predicted long-term outcomes. Moreover, it would be beneficial to also isolate those psychosocial variables that are needed for the pre-trial implant process, before permanent implantation.

Finally, surprisingly, this study found males to be significantly more likely than females to be categorized in the Green and Yellow-I prognosis groups. Because this study did not attempt to elucidate why males are more likely to fall in the better prognostic groups, future research may attempt to compare males and females to determine the differences in biopsychosocial functioning that differentiates them within the prognostic categories. In addition, adverse clinical features were found to be red flags for poor prognostic patients in the present study. Such features have not been extensively studied, and further examination and quantification of these factors in future research would be beneficial. By continuing to elicit data, it will be possible to get a better understanding of how these factors, in sum and individually, affect outcomes. Of course, the literature on long-term studies using PPS is sparse, and there are no current studies taking as many variables into account. Thus,

continuing to look at these data will allow a better look at the long-term efficacy of the algorithm.

In conclusion, the algorithm originally created by Block and colleagues^{5, 14} is potentially applicable to patients undergoing examination for implantable devices to help manage chronic pain. By screening patients prior to implanting SCS and IT pumps, physicians may be able to better choose the patients who will benefit from the invasive and costly procedures. Patients with lower levels of biopsychosocial stress and dysfunction, specifically low levels of depression and more effective coping strategies, appear to be the best candidates for such surgeries. They show the highest percentages of success in overall outcomes, by increasing functional abilities and psychological functioning, decreasing pain levels, and decreasing medication intake. Patients exhibiting large amounts of biopsychosocial stress, specifically high levels of depression and the presence of adverse clinical features, may be poor candidates for these procedures, as they are often unable to recover successfully and tend to have negative outcomes. Targeting those risk factors that appear most indicative of potential success or failure for implantable devices may allow: patients to avoid undergoing procedures that are likely to be unsuccessful; help physicians avoid pitfalls with patients who may not be appropriate for these devices; and create an improved system for third party payers to rely on in order to compensate these costly procedures. Of course, because of the under-powered size of the long-term follow-up sample, ultimate determination of treatment success/failure states of the four groups could not be determined. This awaits future investigation.

Acknowledgments

This research supported in part by grants 5R01 DE10713, 5R01 MH46452 and 1K05 MH71892 from the National Institutes of Health and grant no. DAMD17-03-1-0055 from the Department of Defense

REFERENCES

1. Kumar K, Malik S, Demeria D. Treatment of chronic pain with spinal cord stimulation versus alternative therapies: cost-effectiveness analysis. *Neurosurgery*. 2002; 51(1):106–116. [PubMed: 12182407]
2. Kemler MA, Barendse GAM, Van Kleef M, et al. Spinal cord stimulation in patients with chronic reflex sympathetic dystrophy. *The New England Journal of Medicine*. 2000; 343:618–624. [PubMed: 10965008]
3. North RB, Kidd DH, Zahurak M, James CS, Long DM. Spinal cord stimulation for chronic, intractable pain: Experience over two decades. *Neurosurgery*. 1993; 32(3):384–395. [PubMed: 8455763]
4. Kupers RC, Van den Oever R, Van Houdenhove B, et al. Spinal cord stimulation in Belgium: a nation-wide survey on the incidence, indications, and therapeutic efficacy by the health insurer. *Pain*. 1994; 56(2):211–216. [PubMed: 8008410]
5. Block, AR.; Gatchel, RJ.; Deardorff, W.; Guyer, RD. *The Psychology of Spine Surgery*. American Psychological Association; Washington: 2003.
6. Cashion EL, Lynch WJ. Personality factors and results of lumbar disc surgery. *Neurosurgery*. 1979; 4(2):141–145. [PubMed: 440545]
7. Spengler DM, Ouelette FA, Battie M, Zeh J. Elective discectomy for herniation of a lumbar disc. *J Bone Joint Surg Am*. 1990; 72(2):230–237. [PubMed: 2303509]
8. Junge A, Dvorak J, Ahrens S. Predictors of bad and good outcomes of lumbar disc surgery: a prospective clinical study with recommendations for screening to avoid bad outcomes. *Spine*. 1995; 20(4):460–468. [PubMed: 7747230]
9. Epker J, Block AR. Presurgical psychological screening in back pain patients: A review. *Clin J Pain*. 2001; 17(3):200–205. [PubMed: 11587109]

10. Doleys DM, Klapow JC, Hammer M. Psychological evaluation in spinal cord stimulation therapy. *Pain Reviews*. 1997; 4:189–207.
11. Nelson DV, Kennington M, Novy DM, Squitieri P. Psychological selection criteria for implantable spinal cord stimulators. *Pain Forum*. 1996; 5(2):93–103.
12. Prager J, Jacobs M. Evaluation of patients for implantable pain modalities: Medical and behavioral assessment. *Clin J Pain*. 2001; 17:206–214. [PubMed: 11587110]
13. Block AR. Psychological screening of spine surgery candidates. In: Hochschuler SH.; Cotler, HB.; Guyer, RD., editors. *Rehabilitation of the spine: Science and practice*. Mosby; St. Louis: 1992. p. 617-625.
14. Block AR, Ohnmeiss DO, Guyer RD, Rashbaum RF, Hochschuler SH. The use of presurgical psychological screening to predict the outcome of spine surgery. *The Spine Journal*. 2001; 1:274–282. [PubMed: 14588332]
15. Gatchel, RJ. *Clinical Essentials of Pain Management*. American Psychological Association; Washington, DC: 2005.
16. Anagnostis C, Mayer TG, Gatchel RJ, Proctor T. The Million Visual Analog Scale: Its utility for predicting tertiary rehabilitation outcomes. *Spine*. 2003; 28:1–10.
17. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy*. 1980; 66:271–273. [PubMed: 6450426]
18. Adams LL, Gatchel RJ, Robinson RC, et al. Development of a self-report screening instrument for assessing potential opioid medication misuse in chronic pain patients. *J Pain Symptom Manage*. 2004; 27(5):440–459. [PubMed: 15120773]
19. Beck AT, Ward CH, Mendelson MM, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry*. 1961; 4:561–571. [PubMed: 13688369]
20. Millon, T.; Antoni, M.; Millon, C.; Meagher, S.; Grossman, S. *Millon Behavioral Medicine Diagnostic (MBMD) Manual*. NCS Assessments; Minneapolis: 2001.
21. Butcher, JN.; Dahlstrom, WG.; Graham, JR.; Tellegen, AM.; Kaemmer, B. *MMPI-2: Manual for the administration and scoring*. University of Minnesota Press; Minneapolis, MN: 1989.
22. Ware, JE.; Snow, KK.; Kosinski, M.; Gandek, B. *SF-36 Health Survey: Manual and Interpretation Guide*. The Health Institute, New England Medical Center; Boston: 1993.
23. Rosenstiel A, Keefe F. The use of coping strategies in low back pain patients: Relationship to patient characteristics and current adjustment. *Pain*. 1983; 17:33–44. [PubMed: 6226916]
24. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960; 23:56–62. [PubMed: 14399272]
25. Chapman SL, Jamison RN, Sanders SH. Treatment helpfulness questionnaire: A measure of patient satisfaction with treatment modalities provided in chronic pain management programs. *Pain*. 1998; 68:349–361. [PubMed: 9121824]
26. Greenbough CG, Fraser RD. Comparison of eight psychometric instruments in unselected patients with back pain. *Spine*. 1991; 16:1068–1074. [PubMed: 1835163]
27. Knox BD, Chapman TM. Anterior lumbar interbody fusion for discogram concordant pain. *J Spinal Disord*. 1993; 6(3):242–244. [PubMed: 8347975]
28. Davis RA. A long-term outcome analysis of 984 surgically treated herniated lumbar discs. *J Neurosurg*. 1994; 80:415–421. [PubMed: 8113853]
29. Aicken M, Gensler H. Adjusting for multiple testing when reporting research results: The Bonferroni vs. Holm methods. *Am J Public Health*. 1996; 86:726–728. [PubMed: 8629727]

Figure 1.
Comparison of Prognosis Groups at Initial Evaluation on Oswestry.

Figure 2.
Comparison of Prognostic Groups on MMPI-2 Clinical Scales (Initial Evaluation).

Figure 3.
Adverse Clinical Features of Prognostic Groups at Initial Evaluation (Revised Scoring Methodology).

Figure 4.
Change in Narcotic Use Relative to Prognosis Group.

Table 1A

Statistical comparison of four prognostic groups on demographic variables.

	Green (n=10)	Yellow-I (n=21)	Yellow-II (n=20)	Red (n=9)	χ^2 (df)	p
Gender(%)						
Male	70.0	52.4	30.0	11.1	8.87 (3)	.03
Female	30.0	47.6	70.0	88.9	n.s.	n.s.
Race (%)						
Caucasian	90.0	95.2	85.0	88.9	9.72 (9)	.37
African-Am.	10.0	0.0	10.0	0.0	n.s.	n.s.
Hispanic	0.0	4.8	5.0	0.0	n.s.	n.s.
Other	0.0	0.0	0.0	11.1	n.s.	n.s.
Marital Status (%)						
Married	70.0	76.2	60.0	66.7	7.17(12)	.85
Single	10.0	4.8	5.0	11.1	n.s.	n.s.
Divorced	10.0	4.8	20.0	22.2	n.s.	n.s.
Widowed	10.0	14.3	15.0	0.0	n.s.	n.s.
Disability payments (% receiving)	0.0	38.1	45.0	55.6	7.79 (3)	.05
Pending litigation (%)	0.0	4.8	15.0	11.1	2.51 (3)	.47
SCS implant (%)	50.0	52.4	57.9	33.3	n.s.	n.s.
IT pump (%)	30.0	28.6	21.1	55.6	n.s.	n.s.
Other procedure (%)	20.0	19.0	21.0	11.1	n.s.	n.s.

Table 1B

Statistical comparison of four prognostic groups on demographic variables.

	Green Mean (SD)	Yellow-I Mean (SD)	Yellow-II Mean (SD)	Red Mean (SD)	F(df)	p
Age	60.2 (19.5)	55.5 (16.6)	55.0 (10.7)	51.3 (16.9)	.534 (3,56)	.66

Table 2

Psychosocial Measures of Four Prognostic Groups (Initial Evaluation).

	Green Mean (SD)	Yellow-I Mean (SD)	Yellow-II Mean (SD)	Red Mean (SD)	F(df)	p
BDI	9.30 (4.64)	12.57 (5.57)	17.0 (10.03)	32.22 (15.21)	12.53 (3)	<.001
HAM-D	12.20 (5.25)	12.33 (4.78)	18.35 (5.94)	26.78 (6.46)	17.14 (3)	<.001
CSQ	86.33 (50.0)	99.35 (26.0)	99.0 (42.76)	100.86 (34.68)	.31 (3)	.820
Catast.	8.11 (4.40)	14.37 (7.07)	14.41 (11.02)	27.29 (8.16)	7.06 (3)	.001
MCS	44.75 (7.93)	52.63 (6.80)	46.60 (14.64)	26.0 (13.89)	4.72 (3)	.015

Table 3

Statistical Comparison of Presence/Absence of Adverse Clinical Features

Group	Adverse Features (n)	% Within Group	χ^2 Statistic (df)	p
Green	Absence (9)	90.00	22.76 (3)	<.001**
	Presence (1)	10.00		
Yellow 1	Absence (20)	95.24	4.76	
	Presence (1)	4.76		
Yellow 2	Absence (17)	85.00	15.00	
	Presence (3)	15.00		
Red	Absence (2)	22.22	77.78	
	Presence (7)	77.78		

Table 4

Breakdown of Patients at Six-Month Follow-Up by Prognostic Group

	Total (n=32) n (%)	Green (n=6) n (%)	Yellow-I (n=12) n (%)	Yellow-II (n=10) n (%)	Red (n=4) n (%)
SCS implant (%)	8 (25%)	1 (12.5%)	5 (62.5%)	1 (12.5%)	1 (12.5%)
IT pump (%)	12 (37.5%)	4 (33.3%)	3 (25%)	5 (41.7%)	0 (0%)
Other procedure (%)	2 (6.3%)	0 (0%)	0 (0%)	1 (50%)	1 (50%)
None	10 (31.3%)	1 (10%)	3 (30%)	3 (30%)	3 (30%)