SPINE Volume 24, Number 9, pp 889-898

Restorative Exercise for Clinical Low Back Pain

A Prospective Two-Center Study With 1-Year Follow-Up

Scott Leggett, MS,* Vert Mooney, MD, † Leonard N. Matheson, PhD,† Brian Nelson, MD,§ Ted Dreisinger, PhD,‖ Jill Van Zytveld, BA,† and L. Vie, BA,¶

Study Design. A comparison of treatment of 412 patients with chronic back pain at two centers using the same treatment protocols and outcome measures. Outcome was defined by specific strength testing; Short Form-36 scores at intake, discharge, and 1-year follow-up; self-appraisal of improvement at discharge and in a 1-year follow-up; and reuse of health care services after discharge.

Objective. To investigate the efficacy of standardized treatment methods using isolated lumbar strength testing and strengthening based on progressive protocols using specific equipment. Comparison of results should clarifythe effect of the treatment center versus the efficacy of standardized protocols.

Summary of Background Data. There has been little support in the scientific literature for exercise programs based on standardized protocols. The use of specialized equipment to achieve intense specific exercise also has been poorly supported. Overall health benefit has not often been related to specific improvement in strength.

Methods. More than 400 individuals with chronic back pain were evaluated at the initiation of treatment, discharge, and 1 year after discharge. Measures of efficacy were based on Short Form-36 scores, self-appraisal of improvement, and reuse of health care services after discharge. Study participants were patients with chronic back pain consecutively referred to each treatment site and underwritten by a variety of payers, including workers' compensation, Medicare, and private insurance.

Results. Overall response during the course of the program and at 1-year follow-up was similar between the two centers. Similar proportions of participants at each site demonstrated improvement in SF-36 scores, self-apraisal of improvement, and reuse of health care services.

Conclusions. Standardized protocols using specific strength and measurement equipment can achieve similar benefits at different sites. [Key words: low back pain, outcomes, restorative exercise, strengthening, treatment] Spine 1999;24:889-898

Comparative studies between geographically different centers require that similar protocols and equipment be used. Standardization is critical if the biases of treating clinicians are to be minimized. If identical protocols are used at the beginning, end, and follow-up of such a study, the criticism of a placebo effect also can be negated.

From *U.S. Orthopedics, Little Rock, Arkansas, the †Department of Orthopaedics, University of California, San Diego, California, †Washington University, St. Louis, Missouri, the §Physicians Neck and Back Clinic, Minneapolis, ‖Preventive Care, Columbia, Missouri, and the ¶Columbia Orthopaedic Group, Columbia, Missouri.
Acknowledgment date: April 10,1998.
Acceptance date: August 10, 1998.
Device status category: 1.

Standardization implies that the test methodology will be valid and reliable. Newton and Waddell[24] recently reviewed 10 years of published literature related to the efficacy of so-called dynamic "isomachines." Their report criticized these devices by pointing out lack of standardized protocols and undocumented utility. Further criticism suggested that there was little evidence for reliability of these machines. Although the current authors agree with Newton and Waddell, interestingly the device used in the current article was referenced in support of these criticisms but not discussed in the article. In fact, several studies have reviewed the reliability of full-range isometric lumbar extension testing and dynamic strengthening.[8,9,26,27] These studies document the reliability of this form of testing for healthy individuals and for those with clinical low back pain (CLBP). This form of testing also has both age and gender normative data available and was used in the current study.

Most articles on the topic of CLBP ask the following question: What is the appropriate form of management for CLBP? The assumption made by the authors of this article is that progressive, restorative exercise is the most rational approach. Simple control of back pain by manual therapy and other passive modalities may offer short-term relief, but has not demonstrated long-term resolution of back pain. It is therefore the position of the current authors that reuse of the health care system should be the most important efficacy of a treatment plan.

Although return to work has been used as a definable indicator of a treatment programs success, this is not a reliable outcome measure. This is partly because not all patients with CLBP are in the work force and also because in the worker's compensation system, there are any number of factors that make return to work an inappropriate criterion. Hansen et al[11] documented this in Finland, where exercise intervention for patients with CLBP made no difference in the return-to-work rate. This result was blamed partly on the "off-work rewards" of the Finnish system. Hazard et al[12] found similar lack of correlation between physical measures and return to work.

Lahad et al[16] reviewed a number of articles and found little efficacy for specific exercise programs. Four of these studies suggested that there was no benefit of exercise, whereas seven of the studies noted that increased flexibility and fitness were correlated with decreased low back pain. The other articles reviewed did not discuss the extent of the exercise treatment programs. These studies

were conducted in the workplace and had relatively short follow-up.

Mayer et al[19] documented the strong efficacy of an exercise treatment program with a 1-year follow-up. He found a reuse rate of 30 % in the exercise group. This was half the rate in the nonexercise group. Hazard et al[12] performed a similar study at a different center and found similar results but reported no reuse rates.

The preceding articles were not related directly, and the testing parameters could not be correlated between them because there was no standardization of measures between the articles. Therefore, for comparative purposes, standardization and reuse of the health care system are important.

The purpose of this study was to document the results from treatment of patients with CLBP using standardized testing and training protocols between two centers. Similar progressive, restorative exercise was applied to patients at two centers with patient populations large enough for the use of statistical analysis. The MedX lumbar extension testing and training apparatus (Ocala, FL) was used as an objective measure of low back function and restorative exercise treatment in patients with CLBP.

Another way of comparing different treatment programs is to use a standardized quality of life measurement instrument. Therefore, a further purpose of this study was to compare the results of the Health Status Survey (formerly SF-36) as an instrument of perceived disability.

■ Methods

The current study represents the initial program results and 1-year follow-up of two outpatient chronic low back pain treatment centers. The first center was the Physician's Neck and Back Clinic located in Minneapolis, Minnesota, and the second was the OrthoMed Spine and Joint Conditioning Center in the Department of Orthopaedics at the University of California San Diego (UCSD). Components of the two programs were standardized to include specific high-intensity back strengthening, general strengthening for major muscle groups, 15 to 30 minutes of cardiovascular exercises, and ranging exercises identified by the McKenzie technique. Treatments were provided by physical therapists, clinical exercise physiologists, and certified athletic trainers. Emphasis of treatment was based on muscle function rather than pain symptom reports. A typical program consisted of 2 days per week for 8 weeks on the average. The average cost was $1900 for the UCSD program and a similar figure of $1950 for the Minneapolis program.

Patients were initiated into both programs on the basis of referral from a physician. This study was a prospective study with a 1-year follow-up period beginning after initial discharge. Consecutive patients with CLBP were entered into the program, with no effort made to screen out any patients on the basis of surgery or particular diagnosis. All patients were entered into the study, including those using worker's compensation, private pay, and Medicare.

In all, 1025 patients were entered into the study at UCSD, and 714 patients completed the initial program, as documented by discharge testing. At Minneapolis, 645 patients were entered, with 360 completing the initial program, again as documented by discharge testing. The mean time after onset of current pain was 17 months.

To supply comparable data unbiased by patients with minimal problems, only patients who had completed a minimum of 16 visits were included. Approximately half of the original study group at both centers dropped out before their 16th visit, because they had made sufficient improvement and did not wish to put forth the effort for additional exercise. Approximately one fourth of the dropouts at both centers quit because the exercise was too painful or because they did not like the program. The final one fourth of those who did not complete the program had logistical problems in terms of transportation or insurance coverage, or they had left the area before they could complete the program. These occurrences happened on a random basis.

The patients who completed the program, however, were all included and participated as consecutive patients. A total of 310 patients at UCSD and 102 patients in Minneapolis completed the 1-year follow-up. These became participants for the current study.

At 1-year follow-up, all patients were contacted by telephone for completion of the same study survey that was completed at the beginning of their treatment and at discharge 1 year before. The center at UCSD approached each patient for a physical follow-up to reassess back strength and complete the survey at the clinic. If patients did not make a visit to the center, then a survey was mailed to their home with a self-addressed stamped envelope and returned to the clinic on completion. The Minneapolis center did not attempt to bring patients into the clinic at 1 year after discharge, but pursued all patients by contacting them first and mailing the questionnaires to their homes using the same protocol as that used by UCSD.

Study Survey. The study survey contained initial program, discharge, and follow-up forms. The intake forms contained basic patient demographic and history information, as well as the health status questionnaire (SF-36). The questions related to patient symptoms and history were drawn from a variety of outcome tools and the experience of private practice clinicians.

The SF-36 was part of the study survey and collected on all three occasions: initial, discharge, and 1-year follow-up. The SF-36 questionnaire was developed from the medical outcome survey, which was designed to be a shorts version of standard health perception questionnaires.[21,33,34] The medical outcome survey was a collection of 20 questionnaires selected to represent six areas of health concern. The SF-36 includes these 20 collections and 19 others to represent nine areas of health perception. These tools had been validated on a population of 11,186 patients in a multicenter study to determine the appropriateness of the questions, the ease of independent completion by patient, and the ability to reduce the answers to a few areas of generic health that can be quantified as percentages.[33,34]

The value of the SF-36 is that a patient's changing health perception can be tracked over time to determine the success of rehabilitation and intervention (Table 1). For each dimension, item scores are coded, signed, and transformed into a score from 0 (worst health) to 100 (best health).

Back Strength. Back strength was determined on the MedX back extension machine (Figure 1). This testing device allows for a standardized, isolated strength measurement of the low back extensor musculature. The pelvis is stabilized, allowing

Table 1. Comparison of Subjects Age Between Centers

UCSD		Minneapolis	
Male (n = 131) (yr)	Female (n = 179) (yr)	Male (n = 61) (yr)	Female (n = 41) (yr)
49.3 (17.9)	51.4 (17.0)	39.3 (12.0)	38.7 (12.0)

Values are mean (SD).

no lateral, vertical, or rotational movement, thereby ensuring isolation of the back extensors.

Testing was conducted during the first or second visit to the clinic and at discharge for both UCSD and Minneapolis. It was performed isometrically at Standardized positions of 0,° 12,° 24,° 36,° 48,° 60,° ad 72° of lumbar flexion. Range of motion was determined before the testing, and the closest standardized angle achievable was used for the individual participants. Testing and exercise was performed on the same machine. Initial resistance was set at 50% of maximum isometric torque. Once 15 repetitions could be achieved through the patient's full range at that weight, the resistance was increased by 2% to 5%. For each patient, the treatment program consisted of a workout session twice per week for 8 weeks. Thus, there was an average of 16 specific lumbar extensor strengthening sessions. The training included concentric and eccentric isotonic exercise.

Results

Descriptive Characteristics

Tables 1 and 2 present descriptive characteristics of the 412 participants who completed the 1-year follow-up period in the current study. The diagnostic categories were those assigned by the referring physicians. A two-factor analysis of variance (ANOVA) based on gender and site demonstrated a significant difference in terms of age ($F_{1408} = 36.3$; $P < 0.0001$), showing the UCSD participants to be older than the Minnesota participants. There was no significant between-gender difference, nor an interaction between gender and site. Similarly, there

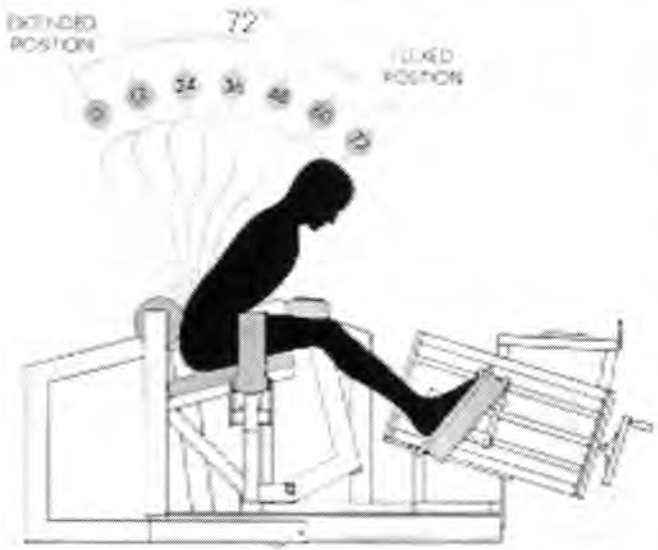

Figure 1. MedX (Ocala, FL) low back machine.

Table 2. Comparison of Subjects' Primary Diagnosis Between Centers

Diagnosis	UCSD (n = 310) (%)	Minneapolis (n = 102) (%)
LB strain/mechanical	65	63
Degenerative disc disease	16	18
Lumbar disc syndrome	11	10
Spondylolisthesis	3	4
Other	4	5

LB = low back.

was no between-site difference in terms of diagnostic category assigned by the referring physician on referral to the program ($P \leq 0.001$).

System reports recorded at intake are presented in Table 3. There were no statistically significant between-site differences in symptom reports ($P \leq 0.01$). Interestingly, of the participants reporting pain in the leg, 62% and 65% reported pain below the knee, respectively for UCSD and Minneapolis.

Regarding previous surgery to the spine, 11% of UCSD participants and 14% of Minneapolis participants reported one previous surgery, whereas 5% of participants at both sites reported more than two previous surgeries

SF-36 Scores

A series of repeated measures ANOVA were performed to examine the effect of treatment site on differences over the three occasions of testing. Treatment site was statistically significant only for the Energy Fatigue variable ($F_{1393} = 5.16$; $P = 0.024$), with Minneapolis participants demonstrating lower intake and follow-up scores than UCSD participants. All of the SF-36 variables demonstrated significant improvement($P < 0.0001$ for all). These results are presented in Table 4.

To consider the effect of gender, a series of repeated measures ANOVA were performed to examine differences over the three occasions of testing. Gender was statistically significant for the Physical Function variable ($F_{1400} = 4.93$; $P = 0.027$) and the Energy Fatigue variable ($F_{1393} = 7.63$; $P - 0.006$). On the former variable, women demonstrated lower intake scores than men. On the Energy Fatigue variable, women demonstrated lower scores than men on all three occasions of testing. As with the ANOVA on site, all of the SF-36 variables demonstrated significant improvement ($P < 0.0001$ for all).
These results are presented in Table 5.

Table 3. Comparison of Subjects' Symptom Reports Between Centers

Site	UCSD (n = 310) (%0	Minneapolis (n = 102) (%)
Back and buttock pain	41	28
Pain into one leg	43	44
Pain into both legs	16	28

Table 4. Comparison of Mean (SD) SF-36 Scores Between UCSD (N = 310) and Minnesota (n = 101) at Each Time of Testing

	Intake		Discharge		Follow-up	
Subtest	UCSD	Minnesota	UCSD	Minnesota	UCSD	Minnesota
General Health	70.5 (22.3)	66.8 (18.5)	73.9 (19.5)	74.2 (17.9)	72.3 (20.6)	66.9 (21.0)
Physical Functioning	60.5 (27.1)	54.0 (22.3)	71.9 (23.2)	74.3 (23.2)	73.1 (24.0)	70.7 (22.8)
Social Functioning	65.9 (29.1)	61.0 (25.6)	81.0 (22.7)	81.8 (22.9)	79.7 (25.7)	76.2 (23.9)
Role Limitation (Physical)	25.6 (37.5)	19.3 (31.4)	53.1 (42.4)	55.9 (42.6)	60.5 (42.5)	50.0 (000)
Role Limitation (Emotional)	72.0 (40.3)	69.3 (40.2)	79.8 (35.2)	86.7 (29.2)	81.7 (33.9)	75.7 (37.0)
Pain	42.0 (22.4)	35.9 (20.0)	62.8 (23.1)	64.4 (23.2)	65.3 (24.5)	60.0 (24.8)
Mental Health	72.7 (18.1)	68.8 (19.5)	77.4 (16.0)	77.5 (16.1)	76.4 (16.6)	72.1 (20.0)
Energy/Fatigue	53.9 (21.0)	46.8 (19.8)	61.4 (18.8)	61.4 (18.1)	60.5 (20.0)	55.1 (21.1)
Health Change	47.9 (21.5)	33.5 (23.4)	60.4 (24.2)	66.7 (23.6)	65.0 (24.2)	67.3 (24.0)

Graphic representations of these results are presented in Figure 2 and 3. It appears that the Minneapolis participants tended to show greater improvement in SF-36 scores at discharge than UCSD participants, although the differences between sites for all measures were not significant.

MexX Back Strength Results

Table 6 presents the MedX average back strength and range of motion results at intake and discharge for each site. There were significant increases in back strength from intake to discharge at UCSD and Minneapolis for flexion (F_{1319} = 389.8; $P < 0.0001$), extension (F_{1322} = 93.6; $P < 0.0001$). There also was a difference in strength improvement between the two programs for both flexion (F_{1319} = 6.14; $P = 0.014$) and extension (F_{1318} = 4.58; P = 0>033), but not for range of motion. These results are presented in Table 6.

No group comparison could be made at 1-year follow-up because Minneaplius did not collect follow-up astrength data.

Overall Appraisal of Successful Outcome

To determine if there were differences between sites in terms of participants' appraisal of outcome, at discharge a chi-square analysis on a rating of "better," "same," or "worse" was conducted. This demonstrated a greater number of "worse: responses for UCSD than for Minnesota (X2 = 10.67; P = 0.0048). Overall, 75% of Minneapolis and 82% of UCSD participants reported that they were better (Table 7).

Regarding the relationship between strength measures and self-appraisal of improvement, there was no difference among the "better," "same," and "worse" groups in terms of extension strength (F_{2315} = 1.29; P = 0.28) or flexion strength (F_{2316} = 0.58; P = 0.56), although there was a difference in terms of range of motion (F_{2319} = 6.03; P = 0.003). Patients describing themselves as "better" over the course of treatment demonstrated greater range of motion at intake. All groups demonstrated similar levels of improvement over the course of treatment (Table 8).

Service Use

To investigate the degree of service use between clinics, an ANOVA on number of treatment visits was conducted. This revealed no difference between the clinics (F1404 = 0.95; P = 0.76), with both averaging approximately 18 visits per study participant. In the year after discharge, 10% of the UCSD participants and 12% of the Minneapolis participants used the health care system.

Discussion

Chronic low back pain is an enormous burden to our society. It accounts for approximately 85% of the total costs for spinal disease, which is estimated to be as high as $80,000 billion per year in the United States. For a problem this pervasive, it is astounding that there is so

Table 5. Comparison of Mean (SD) SF-36 Scores Between Men (n = 191) and Women (n = 220) at Each Time of Testing

	Intake		Discharge		Follow-up	
Subtest	Male	Female	Male	Female	Male	Female
General Health	69.8 (20.2)	69.4 (22.5)	72.9 (19.2)	74.9 (19.0)	69.8 (21.3)	71.9 (20.4)
Physical Function	63.7 (24.8)	54.7 (26.6)	74.0 (22.6)	71.1 (23.7)	73.4 (23.1)	71.8 (24.2)
Social Functioning	64.6 (27.8)	64.7 (28.9)	81.9 (23.1)	80.7 (22.5)	79.1 (25.6)	78.5 (25.5)
Role Limitation (Physical)	25.8 (37.4)	22.4 (34.9)	54.7 (41.3)	53.1 (43.4)	57.2 (43.4)	58.2 (43.4)
Role Limitation (Emotional)	73.4 (39.7)	69.4 (40.7)	85.4 (32.0)	78.3 (35.1)	81.2 (33.5)	79.1 (36.2)
Pain	42.1 (22.1)	39.1 (21.8)	63.3 (23.3)	63.0 (23.0)	64.1 (25.3)	63.9 (24.1)
Mental Health	72.5 (19.1)	71.0 (18.9)	78.5 (14.9)	76.5 (16.8)	75.4 (17.6)	75.3 (17.7)
Energy/Fatigue	55.6 (21.0)	49.0 (20.4)	64.2 (17.6)	59.0 (19.2)	60.0 (20.3)	58.5 (20.5)

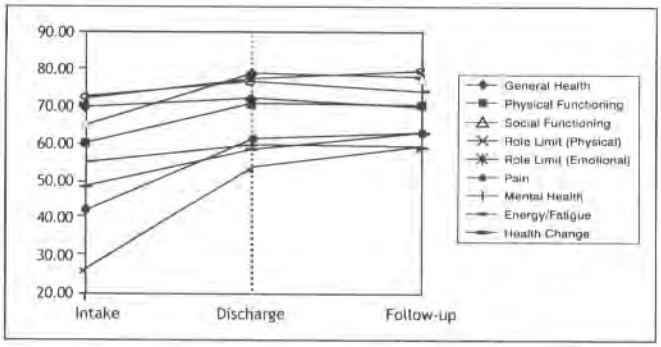

Figure 2. omparison of SF-36 mean scores for all study participants at the University of California, San Diego, at each time of testing.

little agreement about the proper treatment. Because of the extreme reuse of the health care system for CLBP, the current authors strongly believe that reuse is the most important and reliable evaluation of efficacy.

Therefore, the most important statistic to emerge from this study was that showing the reuse of the health care system by patients who had completed a progressive strengthening program. Attempts are always made to prove the efficacy of one treatment plan over another. From a societal standpoint, however, use of resources is the most relevant concern when this benign problem of chronic low back pain is evaluated.

In the current study, there was a reuse rate of 10% and 12%, respectively, of the two centers. The similarity of results documents that this is not an artifact of treater personalities or specific favorable diagnoses. This result occurred just as effectively in workers' compensation situations as in private insurance situations. Is this a significant finding?

A recent study in North Carolina focused specifically on medical use by chronic back patients.[3] In this interview of 4437 adults in North Carolina, 269 individuals reported back pain that had persisted for 3 months or longer and more than 2.5 spells of back pain in the past year. This was extrapolated to indicate an incidence of 3.9%. The use of the health care system was 73.1%, and 10.4% of the patients had undergone surgery.

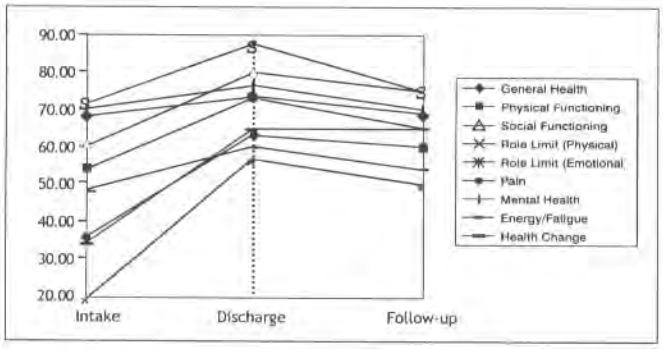

Figure 3. Comparison of mean SF-36 scores for all Minneapolis study participants at each time of testing.

Table 6. Comparison of Mean (SD) MedX Strength Between UCSD (n = 219) and Minnesota (n = 102) at Each Time of Testing

	Intake		Discharge	
Variable	UCSD	Minnesota	UCSD	Minnesota
Extension	80.3 (47.2)	82.1 (53.2)	120.3 (59.2)	146.5 (80.4)
Flexion	142.3 (67.1)	150.3 (73.6)	184.0 (79.8)	220.2 (97.6)
Range of motion	54.5 (16.5)	52.5 (15.7)	61.5 (13.6)	58.7 (11.5)

In a more recent study from the University of Iowa, Lanes et al[17] reported that 55% of those followed up had additional need for medical evaluation.[17] Thirty percent had four or more physician contacts.

Another item of concern in applying a treatment program with a standardized protocol top all patients is the relevance of the underlying diagnosis. It is hoped that the physician's assessment would define the structural source of pain, thus giving some clue as the most rational treatment plan and offering a predication as to resolution. Unfortunately, no study has yet emerged that correlates diagnosis with result. Greeough[10] reviewed the literature, and although noting that psychological aspects may vary prognosis, he found that the medical diagnosis, severity of injury, length of follow-up. Or presence or absence of neurologic deficit made no difference in recovery rates. A more recent study by Michael et al[23] also found a relatively weak agreement between the results of the physical examination and subjective reporting of pain and disability. The current study documented the finding that a defined and focused program specifically designed to restore functional deficits rather than manage pain has an excellent opportunity to provide effective are unrelated to medical diagnosis.

In setting a prospective study, the standard scientific recommendation would be that there should be an unrelated control group for comparison. In the real world of patient care, however, this is extremely difficult. In studies accomplished with no outside funding using funded patients, how could researchers select a group of patients with all the qualifications for appropriate treatment, then inform them that treatment would not be available, but test them the same as the group randomized for treatment? For the size of the studies proposed in this project this was impossible.

Fortunately, a prospective randomized study was conducted by Risch et al[27] using treatment protocols similar to those used in the current study. This group was compared with a nontreatment group. In this study, 54 patients with CLBP were randomly assigned to a 10-week exercise program or placed on wait list to be treated after the control period with the promise of free care. The participants agreed to this randomization before initiation of the program and had the same functional and psychological testing applied before and after the program. The physical training program in this study used

Table 7. Comparison of Reported Appraisal of Treatment Outcome Between Clinic Sites

Appraisa	UCSD	Minnesota
Better	254	75
Same	39	25
Worse	15	1

identical equipment and a protocol very similar to that used in the current study for extensor strength training.

The treatment program was carried out for 10 weeks using specific exercises two times a week for 4 weeks and one time a week for 6 weeks. In addition to lumbar extensor strength testing performed just as in the current study, other psychological and pain perception testing was performed. Specifically, the sickness impact profile was used as one of the main testn tools.[6] The results showed some psychological deterioration in the patients who served as controls. In the control individuals, no change in physical performance occurred after 10 weeks, as expected. In those treated for these 14 sessions, there was significant improvement in strength of lumbar extensors, reduction in pain, and significant psychological improvements.

The patients in the study by Risch et al had CLBP similar to that experienced by patients in the current study. Their average duration of pain was approximately 85 months, and their time off work averaged 2 years or more. This was a study of consecutive patients with no deviation in treatment based on diagnoses, which included postsurgical patients with spinal stenosis and degenerative disc disease. Because the study group in this research project was very similar to those in the current study, the absence of a control group is jot regarded by the current authors as a major restraint in their study.

A recent publication from the University of Iowa emphasized the difficulty in performing long-term follow-up studies.17 In this noncomparative study, 53% of the patients were able to be contacted for a long-term follow-up, which probably represents the best follow-up percentages reported in the literature. Most studies report follow-up percentages of 30% to 40%.

Follow-up study is always is extremely difficult. In the current study, 44% of those who completed the program were followed up at UCSD and 33% at Minneapolis. The authors certainly wished to have a higher number of patients observed at the 1-year review. Their staff made every reasonable effort to pursue the patients. However, both studies were internally funded, and the authors did not feel they could afford additional staff time for extraordinary efforts to search out those who failed to respond to two attempts at follow-up.

Defining outcome presents the most difficult challenge. Unless the study truly is prospective in design, the same evaluation tools cannot be used at the three appropriate points of evaluation: the onset of treatment, the conclusion of treatment, and some long-term specified follow-ups. Definition of change therefore is difficult to assess unless, a specific testing instrument is used. In the current study, strength change was used to evaluate specific lumbar muscle improvement, and SF-36 was used to determine perceived health change, treatment, pain, and other subjective information. There was a 39% increase of strength overall at UCSD and a 43% increase at Minneapolis. Health change as measured by SF-36 changed 30% and 36%, respectively, for UCSD and Minneapolis.

This leads, of course, to the question: What is the appropriate treatment for patients with chronic back pain? The authors of this study make the assumption that the most rational treatment plan for patients with CLBP is a progressive, restorative exercise program. Simple control of pain by manual therapy and various physical therapy modalities may offer short-term benefit, but does not have the opportunity to achieve a long-term reduction in pain complaints. Therefore, probably the most important definition of efficacy is patient reuse of the health care system over a prolonged period. Return to work, for those who are working, is not a valid criteria of efficacy. Although it is a definable end point, return to work is not a reliable outcome measure.

These concepts were documented in an article from Finland reporting on exercise routines used for CLBP treatment of injured workers.22 According to this report, a physical training program made essentially no difference in the return-to-work rate or workplace injury rate of the patients included in the program. The authors accounted for this in terms of the off-work rewards offered by the Finnish system. Significant improvement in physical function was documented. Similar findings showing lack of correlation between physical measures and return to work were noted by Hazard et al.[12] Therefore, it seems important that in addition to reuse of the

Table 8. Comparison of Strength and Range of Motion Scores Across Groups in Terms of Reported Appraisal of Treatment Outcome

	Better		Same		Worse	
Variable	Intake	Discharge	Intake	Discharge	Intake	Discharge
Extension	81.7 (48.2)	131.2 (67.3)	73.5 (50.0)	116.5 (67.3)	97.0 (62.1)	137.4 (82.7)
Flexion	147.2 (70.0)	197.6 (87.0)	134.4 (67.7)	185.8 (95.1)	143.1 (62.2)	195.4 (63.7)
Range of motion	55.3 (15.9)	61.9 (12.3)	49.1 (15.5)	56.2 (14.0)	46.3 (21.0)	55.5 (17.6)

health care system, standardized test instruments focusing on subjective quality of life characteristics should be used in evaluating treatment efficacy.

In addition to the article by Newton and Waddell,[24] others have challenged the utility of exercise programs for the treatment and prevention of low back pain. A recent review article in *JAMA* based on a MEDLINE database search found little documentation for the efficacy of specific exercise programs.[16] The authors pointed out that the evidence was quite skimpy in that all of the articles reviewed were conducted in the workplace.

There was relatively short follow-up, and the total number of participants in the studies was less than 350. In this review, the authors noted that four studies found no beneficial effect of exercises, whereas seven studies found that increased fitness and flexibility was correlated with decreased low back pain. However, of the articles that did not support the finding that exercise is beneficial for patients with CLBP, one was reporting a purely aerobic 8-week exercise program.[4] he other article referenced did not discuss the extent of the exercise treatment programs.

Actually, a few treatment programs have documented their efficacy as related to recurrence rate or other physical and psychological characteristics. The most significant comparable study is that by Mayer et al,[19] which documented the efficacy of a strenuous specific exercise program with a 1-year follow-up. This study had a reuse rate of 30% over the year, which was half that of the comparison group that did not use a progressive exercise program. Mayer et al documented a significant improvement in function in terms of strength and range resulting from an intense 3-week program with 5½-day sessions. In addition to the progressive exercise program, considerable psychological support was provided as well.

The efficacy of str4enuous exercise reported in the preceding study was duplicated by Hazard et al[12] using a similar treatment program with 53 hours of exercise per week, biofeedback, and psychological support therapy. As in the program reported by Mayer et al,[19] specific strengthening exercises for the low back using equipment were used, along with progressive weight training for all the major muscle groups. After the program was completed, an additional 3 weeks of psychological support was provided 1½ to 2 days per week, as well as physical and occupational therapy. Results similar to those found by Mayer et al[19,20] were obtained in terms of positive outcome for 81% of the graduates and very significant improvement in physical function. No data concerning the patients' reuse of the health care system were provided.[12]

The current program was similar to the Mayer and Hazard programs in that it used a comprehensive general strengthening and aerobic program in addition to specific back strengthening exercises on specific equipment. Like the authors of the reported programs, the current authors believed that aerobic training is important for overall metabolic health. Indeed, a warmup before strenuous exercises usually is advised. No injuries occurred in the course of the current treatment program.

There were significant differences as well. The current program did not use psychological support methods and was not as time intensive, although of longer duration. Patients were seen 2 to 3 times a week for 1 to 1½ hours on each occasion, in contrast to a full day of daily exercises for several weeks in the Mayer and Hazard program. At UCSD, McKenzie evaluations were used and home ranging exercises incorporated according to the McKenzie protocols in the treatment plan. The Minneapolis center did not use these evaluations and exercises because the CLBP there was quite chronic. However, this difference apparently was not significant because the efficacy of the programs were quite comparable.

Therefore, one way of simplifying and reducing the overall intensity of the program for CLBP is to focus specifically at the assumed weak link in chronic back disorders. As pointed out by the *JAMA* review article, there seemed to be little scientific evidence showing whether it is a benefit to strengthen flexors or extensors. This may reflect a bias of the current authors in using lumbar extension strength as their outcome measure of muscle treatment because it is true that when the lumbar extensor muscles are compared with flexors by any objective test system, the lumbar extensors are weaker in patients with back pain.

There is evidence in literature from several different evaluation systems showing that in CLBP the lumbar extensors are most likely to be deficient.[31] The justification for the findings of deficient lumbar extensor strength was offered in a study noting metabolic change in the muscle detected by magnetic resonance imaging techniques.[5] This study reevaluated the signal intensity just after exercise and revealed that the multifidus muscles are those most used in lumbar extension. Intensity was much greater in patients with CLBP than in healthy individuals. Abdominal muscles were inconsistently active.

Parkkola et al[25] noted that fatty infiltration in the lumbar extensor of patients with CLBP was higher than that in healthy individuals and higher than that in other muscles. The same findings were noted by Alaranta et al[1] using computed tomography scanning: the more severe the back pain, the greater the atrophy noted in the lumbar extensor muscles only. An even more sophisticated method on analysis is performed with the use of the integrated electromyogram (EMG). Here also, Robinson et al[28,29] demonstrated that the amplitude of the myoelectric signals in the extensors is considerably less than normally expected. This variation in EMG activity compared with that of the lumbar flexors documents the concept that lumbar flexors function normally in patients with low back pain, whereas the lumbar extensors do not.[30]

Sihvonen et al[32] evaluated intervertebral movement and lumbar innervation in 100 patients. These authors found a correlation between radiographic movement of the vertebrae and abnormal medial lumbar EMG activ-

Table 9. Comparison of SF36 Reported Outcomes

Variable	Physical Funtioning	Social Funtioning	Role Limitation (Physical)	Role Limitation (Emotional)	Pain	Mental Health	Energy/Fatigue	General Health Perception
Jenkinson et al: Normayive data for adults of working age								
No long-standing illness	92.5	91.3	91.4	85.6	86.3	75.4	64.0	78.8
Long-standing illness	87.3	80.2	71.9	76.3	69.8	69.9	54.0	60.8
Not seeking current medical attention	89.9	90.5	90.0	85.2	84.6	75.1	63.0	75.7
Seeking current medical attention	81.6	76.9	66.9	72.7	67.7	68.0	52.9	63.7
Brazier et al: Primary care survey of general population								
Healthy	88	89	86	84	82	74	63	73
Seeking current medical attention	81	76	67	73	68	66	52	63
McHorney et al: Primary care survey of general population								
Minor medica	80.5	91.6	70.3	84.3	76.0	82.5	62.0	67.0
Serious medical	57.4	80.0	43.9	76.2	65.1	77.6	47.8	49.1
Kurtin et al: Outpatient dialysis								
Before treatment	47.5	54.7	35.1	34.7	50.3	62.3	34.7	38.1
After treatment	59.5	56.3	28.1	54.1	63.3	66.3	42.6	39.7
Garrett et al: Chronic low back pain, menorrhagia, peptic ulcer,varicose veins (cross sectional comparison; no treatment)		(** $P \leq 0.01$, * $P \leq 0.05$ different from the general population)						
General population	79.2	78.6	76.5	75.0	76.9	73.7	61.2	68.7
Low back pain	52.5*	57.8*	20.5*	45.4*	34.3**	61.5*	42.5*	56.6*
Menorrhagia	77.6	66.5*	46.8*	49.0*	53.4**	59.4*	39.9*	60.0*
Peptic ulcer	80.9	71.6*	60.3*	65.4*	54.0**	64.6*	48.9*	57.2*
Varicose veins	73.9*	78.8	65.4*	66.6*	68.0*	70.6†	56.3*	67.4
Leggett et al: Chronic low back pain at discharge from treatment program								
UCSD	72	81	53	80	63	77	61	74
Minnesota	74	82	56	87	64	78	61	74

* $P \leq 0.01$ vs. general population
† $P \leq 0.05$ vs. general population

ity.Therefore, in the literature, there is considerable evidence showing that it is justifiable to focus the recovery program for chronic pain specifically at lumbar extensors with equipment that adequately isolates this musculature and provides progressive, restorative strengthening exercise.

The validity of SF-36 appears to be well documented. It has been used extensively in a number of different populations. Brazier et al[2] and McHorney et al[21] used the SF-36 with primary care populations to detect differences between patients. Brazier et al compared healthy patients to those defined as "minor medical." McHorney et al compared patients seeking current medical attention defined as "serious medical" (Table 9).

The distribution of scores conformed to what may be expected, thus providing some evidence of construct validity. In the Brazier et al2 study, men had perceived themselves as significantly healthier than women except in general health parameters. Interestingly, there was a significant age variance found for physical function and pain. In other words, physical function decreased and pain increased with age. Health also decreased with lower social class across all dimensions except mental health.

Kurtin et al[15] used the SF-36 in an outpatient dialysis treatment protocol. Interestingly, the study showed changes using the SF-36 in physical functioning and pain, but not in the other SF-36 category parameters. The SF-36 results across the board for the dialysis group were lower than results from other studies, including the current study (Table 9).

The SF-36 also has been reported for functional status outcomes after total knee replacement by Kantz et al[14]. These authors assessed a generic SF-36 response and compared it with an altered SF-36 working on a condition-specific SF-36 response. Results of their study were mixed, but they did find that the generic SF-36 was a good outcome measure for this population as well. Unfortunately, no raw data were presented that could be used for comparison with the current study and previous studies.

The SF-36 data from the current study for UCSD and Minneapolis were consistent with previously reported data for low back pain.[7,13] The current data indicate that patients with CLBP do show a substantial increase in all health categories as measured by the SF-36 outcome tool. After patients completed the standardized progressive exercise program used in the current study, their