

Atlas of standardized low-back tests and measures of the National Institute for Occupational Safety and Health

by Roger M Nelson, PT, PhD, David E Nestor, PT, MS¹

In February 1983 the National Institute for Occupational Safety and Health (NIOSH), Division of Safety Research, initiated a research effort to standardize a set of tests and measures used to assess musculoskeletal injuries to the low back. The initial goal of this NIOSH research project was to provide a reliable, standardized, clinical assessment tool for use by the medical and industrial communities in evaluating low-back musculoskeletal injuries. To accomplish this goal, an expert panel of clinicians, researchers, and educators representing the disciplines of orthopedics, physical therapy, and ergonomics were brought together to develop a methodology for conducting such an ambitious project. During this meeting, a decision was made to concentrate on the largest category of low-back injuries, known as low-back strains and sprains. Since the medical model of disease classification does not adequately describe this category, the panel decided to develop a descriptive schema for low-back strains and sprains by the patients' presenting symptoms, signs, and/or spinal movement dysfunction.

The panel began to assemble a list of tests and measures which had the potential to classify patients according to symptoms and signs. A total of 105 tests and measures were identified that were felt to constitute the "universe" of tests and measurements used during the clinical evaluation of a low-back injury. It was not intended that every patient would be subjected to all 105 test procedures; rather, the collected material simply represented the spectrum of possible tests and measures. The issue of reliability was raised for each test and measure. In the majority of cases, the specific reliability of the test and measure was an unknown factor.

There appeared to be no standardized mechanism by which the tests and measures could be administered with patient-response variables recorded. The initial step toward standardization was to develop a manual describing each test. The individual tests and measures were depicted by a photograph of the procedure, a written description of the clinical procedure, and the potential outcome measures. Upon completion of this initial step, a peer review meeting was held to discuss the collated list.

The review panel recommended that each test and measure be described in a format that eliminated

ambiguity. For example, the anthropometric description of bony landmarks or anatomic sites must follow commonly accepted published material. The refined test procedure was also required to contain a clearly worded description and accurate measurement devices. These steps helped to insure operation procedures for the tests and measures. The reproducibility of the results were enhanced when all the independent variables were standardized. Objective biomechanical standards to measure outcomes were included in the description for each procedure. Twenty tests or measures could not meet these initial criteria even with suggested changes in procedures. They were subsequently eliminated from the original list. The peer review panel recommended that the remaining 85 tests and measures be studied under clinical experimental settings with strict research techniques. The purpose of the study was to estimate the reliability of the individual tests and measures.

The process of conducting one controlled clinical study to evaluate all 85 tests and measures was beyond the scope of one institution. Therefore, the task was broken into five separate categories based on interest and function. The five categories were subjective (history), neurologic, kinesiological, muscle strength/length, and movement testing. Each category had approximately 16 tests or measures assigned to a research site at a major university or clinical center. (This portion of the study is referred to as the "five-site study" in later portions of this report.) The research sites were The George Washington University, the Hayward Orthopedic Center, The University of Iowa, Washington University, and West Virginia University. The advantages of such a division of labor were both rapid turnaround time and broad-based research support. Strict experimental guidelines were developed for the research sites. The research subjects were carefully selected by common criteria to represent those individuals sustaining an initial low-back injury. The patient examiners were trained by the researchers to assure strict adherence to the research protocol. Consistent experimental quality was maintained through the use of an outside group of experimental control advisors. These advisors worked with each site to develop consistent mechanisms of patient selection, test procedure, recording procedure, and quality assurance. To insure uniformity of data collection, the experimental control advisors and the NIOSH project officer reviewed the proposed experimental design. The data

¹ National Institute for Occupational Safety and Health, Morgantown, West Virginia, USA.

response variables were kept to a consistent format. The advisors and the NIOSH project officer met periodically with each site coordinator after data collection had begun. The purposes of the site visits were both to insure adherence to the data collection procedures and to correct minor problems in the testing procedures. As a further quality check, regular telephone consultations were made among the sites, the advisors, and the NIOSH project officer.

This multi-institutional data collection phase involved more than 250 patients with low-back musculoskeletal injuries and 15 examiners. In each test and measure an average of nine subitems were measured, recorded, and subsequently analyzed. These subitems included information on the patient history; low-back range of motion; muscle strength/length; changes in symptoms and signs in response to single, repeated, sustained movements; and alterations in pain/paresis patterns before, during, and after the testing procedures. Data analysis was complicated, producing

more than 12 000 individual statistical analyses. Depending on the scaling type of data collected, a Cohen's kappa coefficient or an intraclass correlation coefficient statistic was used for the data analysis.

At the completion of this data collection phase, a review meeting was held among the individuals responsible for the "five-site study." The reliability of each of the 85 tests and measures was evaluated according to their resultant statistics. A cut-off of 0.75 for Cohen's kappa and 0.80 for the intraclass correlation coefficient was used to assign the test or measure to the "acceptable" category. Those tests and measures with Cohen's kappa and intraclass correlation coefficients that did not reach the acceptable cut-off, but were deemed useful by the clinical experts and were felt to have potential for improvement, were assigned to the "marginal" category. Those tests and measures which had a Cohen's kappa and/or an intraclass correlation coefficient below 0.40 and did not meet the "marginal" criteria were assigned to the "low" cate-

Table 1. Low-back atlas of standardized tests and measures.

Starting position	Test/measure	Description
Standing		
1	Lumbar lordosis	Extent of the lordotic curve measured with a flexible ruler
2	Iliac crest (posterior)	Level of the crest estimated by the crest tester and visual assessment
3	Posterior superior iliac spine (right/left)	Level of the posterior superior iliac spine estimated by the crest tester and visual assessment
4	Iliac crest (anterior)	Level of the crest estimated by the crest tester and visual assessment
5	Anterior superior iliac spine (right/left)	Level of the anterior superior iliac spine estimated by the crest tester and visual assessment
6/7	Side bending to the right and to the left	Excursion (range of motion) of side bending to the right and then to the left measured with a vertical ruler
Sitting		
1/2	Iliac crest (anterior and posterior)	Level of the crest estimated by the crest tester and visual assessment
3	Posterior superior iliac spines (static)	Levels of the posterior superior iliac spines at rest measured by visual assessment
4	Lumbar lordosis	Extent of the lordotic curve measured with the subject in a relaxed sitting position
5	Lumbar lordosis during sustained slouched sitting	Extent of the lordotic curve measured with a flexible ruler with the subject in a slouched seated posture, which is an indirect measurement of flexion
6	Lumbar lordosis during sustained erect sitting	Extent of the lordotic curve measured with a flexible ruler with the subject in a sustained erect sitting posture
Prone		
1	Hip range of motion (internal/external)	Degrees of available motion in both hips, internal and external rotation, measured with a goniometer
2	Sustained extension in lying (press up)	Assessment of the change in symptoms
Supine		
1	Straight-leg raise	Angle of the straight-leg raise measured with a bubble (gravity) goniometer; no reproduction of symptoms
2	Double straight-leg raise	Angle of the double straight-leg raise measured with a bubble (gravity) goniometer
3	Hamstring length	Relative length of the hamstring muscles measured with a bubble (gravity) goniometer
Side-lying		
1	Gluteus medius strength	Assessment of muscular strength of the gluteus medius muscle

gory. Some tests and measures could not be tested because they had too few subjects in that category or the distribution of subjects was not sufficient in that category for analysis. These tests and measures were put into an "insufficient distribution" category.

Sixty-eight "acceptable" or "marginal" tests and measures were assembled to be examined in a clinical trial study. The marginal tests were included in an attempt to ascertain whether their reliability could be improved by revision of specific areas felt to be weak by the examiners. Knowledge gained from the five-site study was applied to refine each test and measure for the clinical trial study of the 68 tests and measures. Each test and measure was restructured and standardized according to its test description, the test instrument required, definition of terms, contraindications, starting position of the patient, starting position of the examiner, procedure, order of assessments, timing, rest periods, procedure if the patient could attain or maintain position, indications for discontinuing the test, position in which to leave the patient, and common errors to avoid. Since the reliability estimates would be carried out in a clinical setting with patients with low-back injuries, the rest periods and test/measure ordering was of prime importance. Many iterations of the ordering of the tests and measures were performed before the final list was established.

The clinically based study at Washington University was conducted with a total of 53 subjects. The

analysis for each reliability statistic, depending on the scaling type of data, was the same as that of the five-site study mentioned earlier in this report.

Upon the completion of these clinical trials, a meeting was held with the contractors, the clinical experts who participated in the trials, and the NIOSH project officer to review each test and measure for reliability. In addition, each test and measure was classified into one of the following four categories: acceptable, marginal, low, and insufficient distribution. As a result of this meeting, a final list of 19 acceptable and marginal tests and measures has been compiled (table 1). This list was reviewed by a panel of eminent health care practitioners known for their expertise in evaluating and treating patients with low-back injuries. The peer review comments and suggestions have been incorporated into the compiled atlas.

In summary, this historical account presents a model for the development of a standardized list of tests and measures which may be used to assess a low-back injury. The methodology employed to develop the low-back atlas may also serve as a pattern for the development of similar assessment systems for upper and lower extremity musculoskeletal injuries. The future goals for the Division of Safety Research will be to (i) develop a classification system for patients with low-back musculoskeletal injuries, (ii) develop a clinical low-back assessment system, and (iii) develop intervention and prevention strategies.