Interexaminer agreement on palpatory diagnosis and patient self-assessment of disability: A pilot study

MYRON C. BEAL, DO
DAVID A. PATRIQUIN, DO

To evaluate interexaminer agreement on palpatory diagnostic findings, we compared interexaminer results, patient disability self-evaluations, and assessment of patients' progress by referring physicians. Three examiners (two clinicians and a third-year osteopathic medical student fellow) monitored patients' progress using negotiated examination procedures. The patients were examined independently by each examiner at each of eight weekly visits. Patients completed a disability assessment form at each visit, and the referring physicians recorded changes in their patients' condition. The patients' disability self-rating and examiner test results did not show clear correlation. There was a 62.2% agreement between the two physician examiners when the general descriptors "improvement," "no change," or "aggravation" were used. The student examiner's agreement with the clinician examiners was 60.2% and 51.8%. Interexaminer agreement of findings from osteopathic testing procedures appears to depend on general clinical experience and specific experience with the testing procedures.

(Key words: Interexaminer agreement, osteopathic palpatory findings, disability self-assessment)

During the past 40 years, interexaminer reliability studies[1-15] have been conducted employing two or more independent examiners who have carried out a structural examination identifying sites of somatic dysfunction or using a limited protocol of specific palpatory tests. Some of the objectives of these studies were to correlate examiner results, examine the reliability of the examination method and of the specific tests, develop accurate criteria for palpatory tests, analyze findings of disagreement, look at the training for interexaminer testing, compare the test results of students and trained examiners, study patient improvement, and correlate palpatory findings with patient diagnosis. These studies have also employed variations in subject and examiner blinding.

In a previous research study, a group of negotiated tests was used to monitor the patients' progress during treatment.[13] We made changes and additions to this study, which included the following:

- The examiners included a student fellow, who was a predoctoral osteopathic medical student committed to an extra (fifth) year at the college, during which time he assisted in the teaching, treating, and research programs of the osteopathic principles and practice faculty. His diagnostic and treatment skills were subject to evaluation as part of this study.
- Each examiner independently documented the complaint of the patient at the initial interview. The examiner selected palpatory tests appropriate for the evaluation of the patient.
- The study included patients currently receiving osteopathic manipulative treatment (OMT) and a control group consisting of patients who were not receiving OMT.
- The examiners were blinded to the subjects' treatment program.
- Examiners were permitted to review their previous test records before each weekly reexamination.
- Each patient completed a disability assessment form at each visit, with the assistance of a second student fellow.
- The referring physician was asked to record both his examination findings and his evaluation of the patient at each visit.

From the Department of Family Medicine, Osteopathic Manipulative Medicine, Ohio University College of Osteopathic Medicine, Athens, Ohio (Dr Patriquin). Dr Beal is Emeritus Professor, Michigan State University College of Osteopathic Medicine, East Lansing, Mich. Dr Patriquin is currently semiretired.

Correspondence to David A. Patriquin, DO, Ohio University College of Osteopathic Medicine, Grosvenor Hall, Athens, OH 45701.

PATIENT DISABILITY ASSESSMENT

Clinical Visit

Doe	John	X	01-01-99
Last	First	M	Date

Check activities according to your present level of performance as either normal or slightly moderately, or severely impaired.

Factors to consider are quickness of action, coordination of movements, strength, security, endurance, pain, concern, worry.

Please rate each activity with a mark at the appropriate place on the scale

ACTIVITY	SCALE
1. Work: are you able to carry out your normal activities?	NORMAL Slight Moderate Severe
2. Recreation: are you able to pursue hobbies, sports, leisure activities?	NORMAL Slight Moderate Severe
3. Physical suffering: are you free from malaise, pain, or suffering?	NORMAL Slight Moderate Severe
4. Mental suffering: are you free from worry or unhappiness?	NORMAL Slight Moderate Severe
5. Sleep: is your sleep satisfactory?	NORMAL Slight Moderate Severe
6. Depending on others: are you independent of others for acts of daily living (washing, feeding, dressing, moving)?	NORMAL Slight Moderate Severe
7. Feeding: are you eating, do you enjoy your food?	NORMAL Slight Moderate Severe

X = 1st week's self-assessment

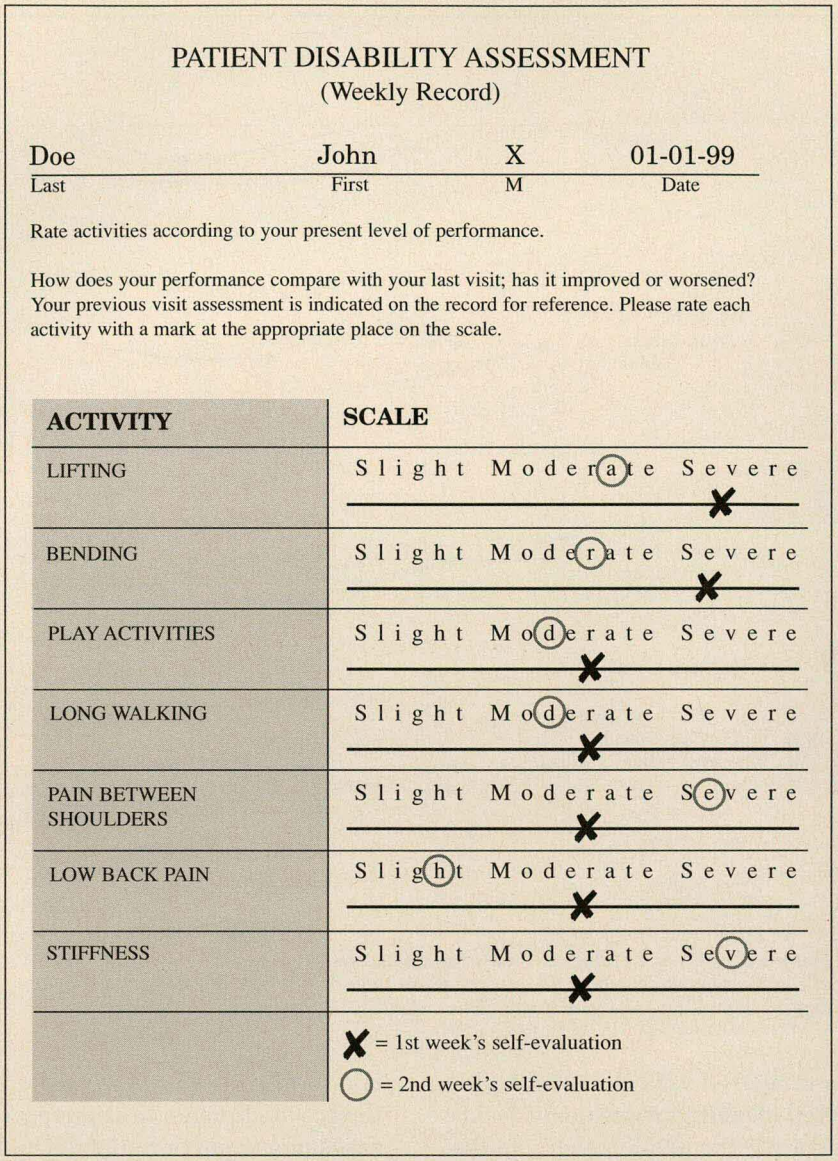
Figure 1. *Patient disability assessment form (initial visit).*

The goal of this study was to evaluate the interexaminer agreement on palpatory diagnosis. We compared interexaminer test results, patient disability self-evaluation forms, and the assessment of patients' progress by the referring physician.

Methods

The subjects for this study were referred by physicians in the ambulatory clinics at the Ohio University College of Osteopathic Medicine (OU-COM). Two groups of patients were sought for this study: those receiving OMT as a part of their treatment program, and those not receiving OMT. A research coordinator selected subjects for the study. Subjects were selected who were adults and who were available weekly for 8 consecutive weeks. The research protocol called for 10 patients who were receiving OMT and 10 patients who were not receiving OMT.

Osteopathic palpatory examinations were conducted by two osteopathic physicians and a senior student fellow. The examiners were unacquainted with the patients' treatment program. At the initial visit, each examiner independently obtained a brief report of the patient's presenting complaint and conducted a structural examination. Examination tests and their results were recorded. After the three examiners had completed their separate evaluations of the patient, they met together with the patient and reviewed their respective tests and results. Tests to



PATIENT DISABILITY ASSESSMENT
(Weekly Record)

Doe	John	X	01-01-99
Last	First	M	Date

Rate activities according to your present level of performance.

How does your performance compare with your last visit; has it improved or worsened? Your previous visit assessment is indicated on the record for reference. Please rate each activity with a mark at the appropriate place on the scale.

ACTIVITY	SCALE
LIFTING	Slight Moderate Severe
BENDING	Slight Moderate Severe
PLAY ACTIVITIES	Slight Moderate Severe
LONG WALKING	Slight Moderate Severe
PAIN BETWEEN SHOULDERS	Slight Moderate Severe
LOW BACK PAIN	Slight Moderate Severe
STIFFNESS	Slight Moderate Severe
	✗ = 1st week's self-evaluation ○ = 2nd week's self-evaluation

Figure 2. *Patient disability assessment form (subsequent visit)*

be used to monitor the patient's progress then were selected by the examiners. Tests were selected from the following classes identified by Dinnar[16]: class III, palpatory definition of landmarks; class IV, palpation of tissue texture, both superficial and deep; and class V, specific joint motion testing.

The initial findings were assigned a value of *0* regardless of their intensity. The full scale was −4 to +4. On subsequent evaluation, improvement was recorded as a minus increment, aggravation as a plus increment, and no change as the same scale value as the preceding week's score.[3] Scale values of −3 and +4 indicate minimal intensity or maximal severity of examination findings.

At the initial visit, another senior student fellow assisted the subject in the completion of an Initial Patient Disability Assessment form (*Figure 1*). General and specific complaints were identified and recorded with their intensity graded on a 20-point scale from "slight" to "severe." The "menu" items on the initial form focused the subject's complaints. From these general subjects, the second student fellow identified specific (component) complaints, which were transferred to a weekly record (*Figure 2*). At each of the eight weekly visits, the student fellow would complete, with the subject, a new weekly disability assessment form. The form would be marked in advance by the second fellow with the scores assigned by the patient at the previous visit. Thus, the patient, using the 20-point scale, indicated the cur-

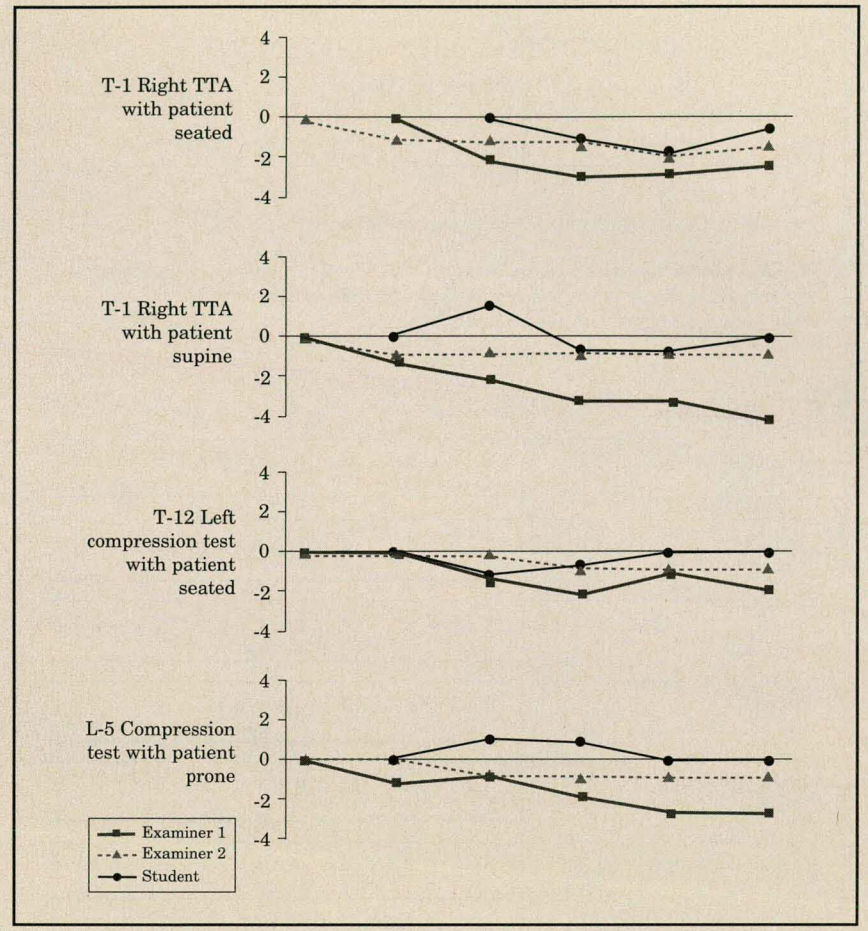


Figure 3. *Patient 2. Judgments of three examiners on palpatory diagnostic procedures at eight weekly visits. On scale, minus values show improvement, and plus values show aggravation. TTA indicates tissue texture abnormality palpated in the soft tissues.*

rent disability score compared with that of the previous week. Completed disability assessment forms were secured in a separate file available only to the second student fellow.

At each visit, the three examiners would evaluate the patient independently and in random order, using only the specific tests agreed to for that subject. Examiners were encouraged to refer to their record of previous visits before and during each examination and to avoid discussing the subject's clinical condition or to engage in any treatment.

The random ordering of examiners was used to lessen the bias arising from the changes in structural findings that might result from serial physical examinations. Stable findings were selected for test evaluation. The character of findings to be included and excluded in the study was not stated in advance. It was agreed that structural findings to be included in the testing protocol should be closely related to the principal complaints of the patient.

Referring (source) physicians were asked to maintain specific progress notes on their patients. They were to record changes in their patients, status in the assessment portion of their subjective, objective, assessment, and plans (SOAP)–format progress notes.

Results

Of the 10 patients included in this pilot study, 8 were receiving OMT and 2 were not. The 5 men and 5 women ranged in age from 23 to 63 years. There were not enough subjects to match the treated and untreated groups as to gender and age.

Statistical evaluation of agreement between examiner pairs, based on numerical evaluations (−4 to +4) assigned by the examiners, showed the following mean percentage agreement: between examiner 1 and examiner 2, 38.9%; between examiner 1 and the student, 39.5%; and between examiner 2 and the student, 40.4%.

Statistical evaluation of agreement between

(continued on page 103)

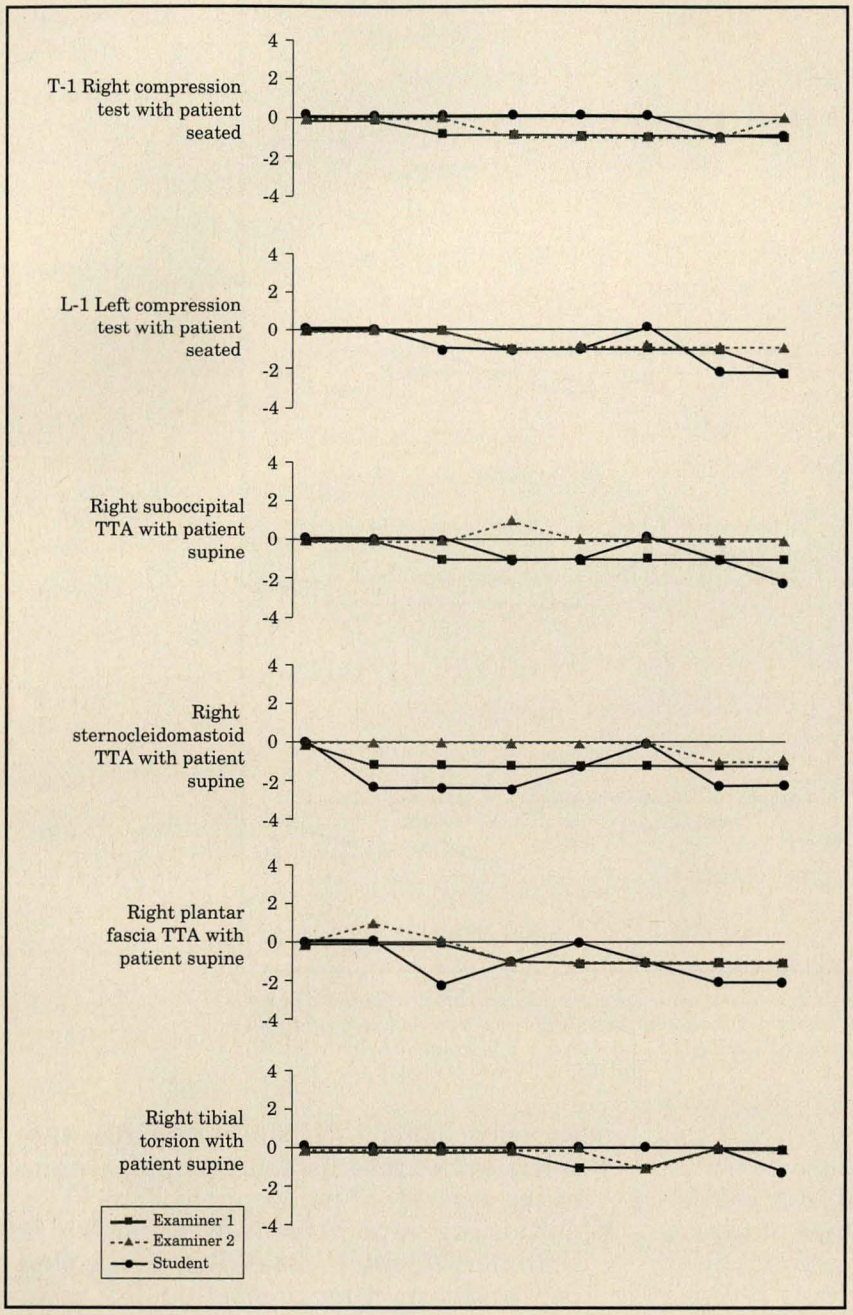


Figure 4. *Patient 8. Judgments of three examiners on palpatory diagnostic procedures at eight weekly visits. On scale, minus values show improvement, and plus values show aggravation. TTA indicates tissue texture abnormality palpated in the soft tissues.*

examiner pairs, based on nonnumerical test results (improvement, no change, or aggravation) assigned by the examiners, showed the following mean percentage agreement: between examiner 1 and examiner 2, 62.2%; between examiner 1 and the student, 60.2%; and between examiner 2 and the student, 51.8%. Each examiner was relatively consistent in his own ratings. Also, the comparative evaluation reports of two clinicians tended to be more consistent than were those of the student fellow with either of the clinicians. However, a 10% gap between student/clinician and clinician/clinician figures was translated as an acceptable level for defining agreement.

At each subject's visit, an average of four to six tests were performed by each examiner. There was agreement between the physician examiners in 50% or more of the tests performed on eight of the subjects. The subjects on whom there was less than 50% agreement were those not receiving OMT. The physician examiners and the student fellow agreed in 50% or more of the tests performed on five of the ten subjects.

Examples of interexaminer agreement are displayed graphically in *Figures 3* through *5*. Agreement is indicated by corresponding trends in plotted examiner evaluations.

We did not demonstrate a clear correlation between examiner test results and the patient disability assessment self-evaluation. In four cases, patient-reported improvement and examiner-reported improvement agreed. Three subjects reported worsening of their disability in the presence of examiner reports indicating improvement. Two subjects who re- ported aggravation of their complaints were found by the examiners to have improved in less than half of their tests; one patient who reported improvement was found by the examiners to have improved in only one test.

The progress notes of the referring physicians were inadequate for determining whether the subjects' complaints had changed. In two of the three cases in which the referring physician noted patient improvement, the patients' disability self-evaluation disagreed.

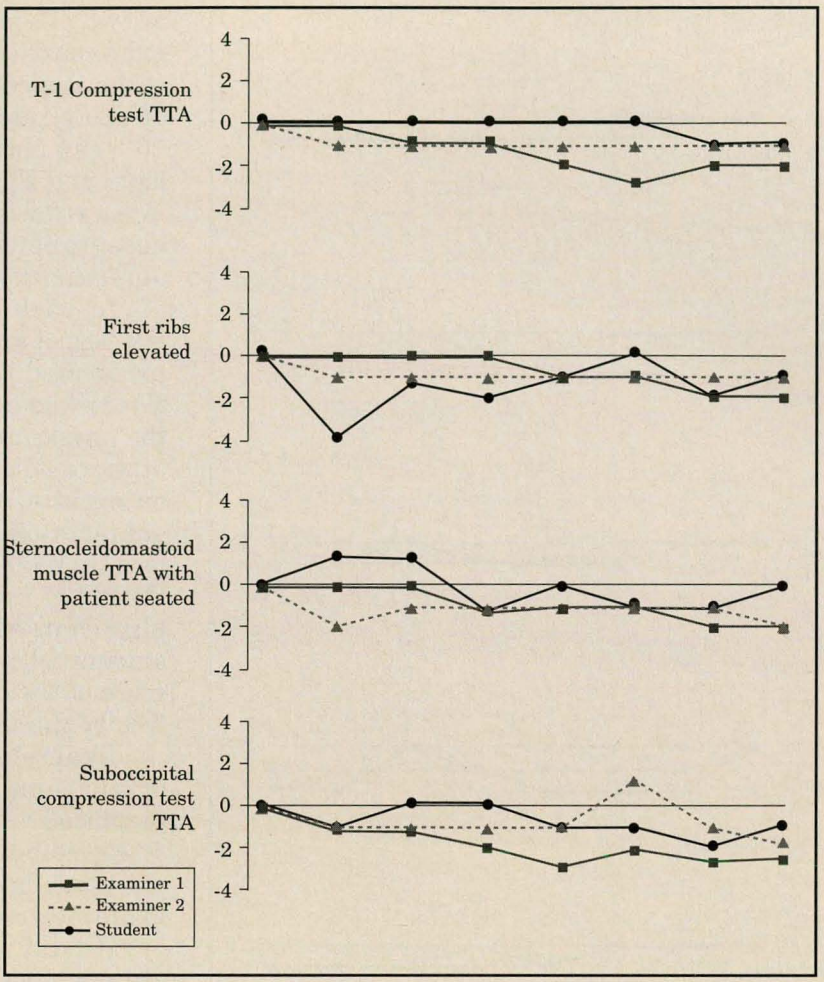


Figure 5. *Patient 9. Judgments of three examiners on palpatory diagnostic procedures at eight weekly visits. On scale, minus values show improvement, and plus values show aggravation. TTA indicates tissue texture abnormality palpated in the soft tissues.*

Discussion

The protocol for this study includes several changes and additions suggested in a previous research report.[13] The correlation of these sets of data, examiner test results, patient disability self-evaluation, and referring physician progress notations increased the complexity of the design and process of this study.

Preparation required coordination and cooperation of many individuals and adequate preparation and understanding of their respective roles. Some of the problems encountered were in the following areas:

- Training the personnel in their respective roles, and indicating their importance to the success of the project. This group included clinicians, the research coordinator, nursing personnel, and appointment clerks.
- Obtaining subjects from clinicians who had never participated in a clinical research study.
- Adequacy of records and the maintenance of the security of independent data files.
- Obtaining appointment of a visiting researcher (Myron C. Beal, DO) to the clinical faculty in order to meet quality assurance and malpractice insurance requirements in the clinic.

A shortened subject recruitment time coupled with inexperienced on-site personnel strongly contributed to the small number of subjects in both treated and untreated groups. This small study sample prevented an adequate comparison of the examiner agreement in respect to patient progress in these two groups of subjects. Some patients were not examined each week because of scheduling difficulties.

The referring physician's role was critical both in the proposing of patients for the study

and in a careful documentation of evaluation and treatment in the progress notes. The role of the referring physician was not adequately explained to the participating clinicians nor were they given adequate training.

The tests were negotiated by the three examiners after each had independently examined the patient at the initial visit. Tests were described by patient position, location of the area or structure to be examined, asymmetry of position, tissue texture changes or abnormality, and joint mobility alteration. Uniform test description facilitated the examiner's location of the site to be tested. In some cases, however, two examiners agreed to use a test preferred by the third examiner even though they were not comfortable using it. No time was provided for test practice to develop research level skills in application. All examiners should be competent in the use of all tests adopted for a research project that depends on data from palpatory evaluation of the musculoskeletal system. More practice, both in the initial phase and at regular intervals during the 8 weeks of the study, would have improved the quality of the study.

The tests selected by the examiners were related to the subject's history, but were not necessarily focused on areas currently being treated by the referring physician. The examiners were unaware of the treatment plan of each subject. The progress notes of the referring physicians, whose patients were receiving manipulative treatment, indicated that most areas studied in the research project were treated by that referring physician. The notes of the referring physicians indicated more general areas of treatment than the specific joint and muscle groups identified by the research examiners. In some instances, test sites were not treated by the referring physician. An example was a contraction in the superior gluteal musculature associated with low back pain. The two clinicians found no change in the test of the gluteal muscle contraction during the 8-week study, whereas the student examiner found improvement in this patient in the sixth week.

Even though examiners had access to their record when examining a subject, it became apparent that exact recall of the results of the previous week's testing was of paramount importance in the evaluation. The examiners agree that memory of palpatory findings is reinforced in a normal office visit where more time is taken to examine, treat, and reexamine the patient. The repetition of examination after treatment makes an important contribution to the examiner's memory of palpatory findings at subsequent patient visits.

Comments

The examiners found improvement in four subjects who reported improvement on their disability assessment forms. The examiners also cited improvement in three subjects who reported their condition to be worse. The noncorrelation between examiner test results and subject disability self-assessment underscores the subjective nature of both processes. Psychologic and attitudinal factors in the subjects affected these data. For example, one of the subjects, who was in the midst of her university final examinations, scored all her disability categories worse that week.

When the student fellow's data were compared with those of one of the examiners, there was agreement in eight of the ten patients. He agreed with both of the examiners in 50% or more of his tests on five patients (50% of the patients). It is noteworthy that a fourth-year osteopathic medical student has achieved this level of palpatory skill.

This type of research has an intrinsic weakness in that it depends on data of subjective origin. Reliable, research-quality interexaminer agreement of palpatory test results depends on adequate preparation of the examiners. Training to develop skill in palpation and consistency in the process of examination is imperative to identify findings and changes in them over time. Everything possible must be done to establish the skill limits of the examiners. The tests selected should not require greater ability than that of the least qualified of the examiners. Time should be set aside before the study begins to evaluate the abilities of the examiners, to identify the tests to be used in the study, and to practice the exact protocol for test procedures. Test practice sessions at stated intervals after the project has started should improve interexaminer conformity in the testing procedures. The review should include process, interpretation, and recording of each test.

Interexaminer research studies involving palpatory examination of the musculoskeletal system depend on data derived from subjec-

tive evaluation processes. Studies[6] have shown that interexaminer reliability can be improved by careful selection and training of examiners and complete, clear descriptions of test procedures.

Study of the testing process gives rise to several questions. Do all tests produce data of equal value? Do some tests permit more useful correlation with the progress of the subject? Do palpatory tests correlate with objective measurement of the subject's condition? Can we identify the range of findings to be expected on applying a specific palpatory test? Answers to these, and many related questions, are critical to the development of reliable clinical research in osteopathic palpatory diagnosis.

The study showed that two experienced osteopathic physicians can achieve a 62.2% level of agreement on palpatory diagnostic findings. The osteopathic medical student's level of agreement is comparable to that found in current research on palpations.

This research protocol was designed to compare the interexaminer agreement of two experienced osteopathic physicians using negotiated testing procedures to determine patient progress, to compare the agreement of a student fellow with experienced clinicians, to compare the patient's disability assessment with the examiners' evaluation of the patient's clinical progress, and to compare the referring physicians' evaluation of the patient with the patients' disability assessment and the examiners' evaluation of patient progress. The results of such studies should be of benefit to the osteopathic teaching of palpatory and treatment skills.

Acknowledgment

Our appreciation goes to many people who assisted in this research study: Ms Elizabeth Randolph, for statistical help; Daniel Cain, DO, and James Muntean, DO, student fellows in osteopathic principles at the time this study was conducted; OU-COM Clinic Research Coordinator Jean Moquin, RN; referring physicians Anthony Chila, DO, John Glover, DO, and Daniel Marazon, DO; Mr Jeff Brown and Ms Danette Pratt for graphics.

References

1. Denslow JS: Symposium: An approach to skeletal components in health and disease. *JAOA* 1951;50:399-403.

2. McConnell DG, Beal MC, Dinnar U, et al: Low agreement of findings in neuromusculoskeletal examinations by a group of osteopathic physicians using their own procedures. *JAOA* 1980;79:441-450.

3. Kappler RE: A comparison of structural examination findings obtained by experienced physician examiners and student examiners on hospital patients. *JAOA* 1980;79:468-471.

4. Dvorak J, Dvorak V, Schneider W: Controlled patient examination trial. *Manual Med* 1985;1:100:115.

5. Johnston WL, Elkis ML, Marino RV, et al: Passive gross motion testing: Part II. A study of interexaminer agreement. *JAOA* 1982;81:304-308.

6. Johnston WL, Beal MC, Blum GA, et al: Passive gross motion testing: Part III. Examiner agreement on selected subjects. *JAOA 1982;81:309-313.*

7. Johnston WL, Hill JL, Sealey JW, et al: Palpatory findings in the cervicothoracic region: Variations in normotensive and hypertensive subjects. A preliminary report. *JAOA* 1980;79:300-308.

8. Johnston WL, Allan BR, Hendra JL, et al: Multiple examiner study of spinal segmental dysfunctions, abstracted. *JAOA* 1981;80:752-753.

9. Upledger JE: The reproducibility of craniosacral examination findings: A statistical analysis. *JAOA* 1977;76:890-899.

10. Gonnella C, Paris SV, Kutner M: Reliability in evaluating passive intervertebral motion. *Phys Ther* 1982;62:436-444.

11. Kaltenborn F, Lindahl O: Reproducibility of the results of manual mobility testing of specific intervertebral motion. *Swed Med J* 1969;66:962-965.

12. Beal MC, Goodridge JP, Johnston WL, et al: Interexaminer agreement on patient improvement after negotiated selection of tests. *JAOA* 1980;79:432-440.

13. Beal MC, Goodridge JP, Johnston WH, et al: Interexaminer agreement on long-term patient improvement: An exercise in research design. *JAOA* 1982;81:322-328.

14. Beal MC, Dvorak J: Palpatory examination of the spine: A comparison of the results of two methods and their relationship to visceral disease. *Manual Med* 1984;1:25-32.

15. Cox JM, Gorbis S, Dick LM, et al: Palpable musculoskeletal findings in coronary artery disease: Results of a double-blind study. *JAOA* 1983;82:832-836.

16. Dinnar U, Beal MC, Goodridge JP, et al: Classification of diagnostic tests used with osteopathic manipulation. *JAOA* 1980;79:451-455.