Interexaminer reliability studies: Spanning a gap in medical research—Louisa Burns Memorial Lecture

WILLIAM L. JOHNSTON, D.O., FAAO
East Lansing, Michigan

In an emerging clinical science such as osteopathic medicine, a clinician's descriptive record provides the nonphysician scientist with questions to be answered through basic and applied research. In this address, progress made in documenting osteopathic palpatory test procedures during the past half century is reviewed. During this period of clinical development, it has been important to recognize that within the physical examination there is a logical progression in testing for signs, in identifying location, and in carefully defining a problem, sufficient to develop a focus for treatment. The same logical sequence is seen for diagnostic tests of the somatic system that detect not only the presence of regional somatic dysfunction, but also the specific location and definition of problem areas of segmental dysfunction for their consideration in patient care. Results of multiple examiner studies of palpatory tests from each level in the diagnostic sequence are reviewed. These studies provide an essential link between clinical observation and rigorous research, stimulating the development of further directions of applied and fundamental human research. With improved clinical and instrumental tests for diagnosis of the somatic system, it is now feasible to project controlled studies in a clinical setting with manipulative treatment as the experimental variable.

It is indeed a singular privilege to be given the opportunity of addressing researchers in the osteopathic profession on this occasion of honoring our pioneer researcher, Dr. Louisa Burns. As a clinician standing before a group of scientists, I feel a certain kinship to Northrop Frye,1 also Canadian-born, and a humanist, speaking before a group of scientists with this introduction: "My chief qualification for addressing you here is my total ignorance of everything you know." Dr. Frye proceeded to draw on his experience in the humanities in building a language bridge between their two "cultures," and if I draw on my own clinical experience during this address, I hope it may serve as a bridge between the two cultures in this audience, D.O. and Ph.D.

Dr. Louisa Burns2 conducted her experimental research in close relation to her own personal clinical experience. Even today, this is a rare combination—the laboratory researcher who is also an active clinician. In his book, Clinical judgment, Feinstein assures us that "During a single week of active practice, a busy clinician conducts more experiments than most of his laboratory colleagues do in a year."3 You may find the point controversial. You may agree, however, that the scientific basis of medical practice is not just applying what is scientific fact to a clinical situation. It is the attitude of the physician who realizes that every clinical situation is a scientific investigation, ripe for careful observation of the patient, and opportune for repeatedly testing efficacy of treatment. Because the situation of the experimenter who is a clinician is rare, the experimenter often looks to clinical observation and communication with the clinician for questions that arise in medical practice and provide a practical focus for fundamental research. The clinician in turn brings the recognition of the unusual clinical experience to the basic scientist for scholarly consideration and technical study.

I would like to talk with you about the role of interexaminer studies as one link (and an important one) in this communication between the clinician
and the basic scientist who are each devoted to medical research. I will be touching on 50 years of osteopathic descriptive research that have made interexaminer studies possible, and previewing directions of research that have been created through our studies.

First of all, what is the area of medical research that has special meaning for the osteopathic profession? Let me give you an example, not with the intention of delimiting the area, but rather to develop a perspective as well as a focus for my remarks. I am walking behind five grade school students, on a street in Chicago's own Hyde Park, to be specific. They are not a gang; they're an average group of boys aged 9 to 12 years, on their way home after school. The smallest is thin, wiry, straight and distinctively muscular frame. The fifth boy is sinewy, crouch to the shoulder carriage and the head as a ramrod, and leading the group with the quickest strut. The second boy has an unloaded toy BB gun; he's cocking it and swinging it, and talking a blue-streak. The two tallest in the center are similar in athletic build, but one has a more limited and asymmetric, and how this bears on his future health, his resistance to respiratory infections, or his predisposition to cardiovascular disease, or possibly his future as a hypertensive patient.

In contrast, the body types I was observing have been a subject for classic study by Sheldon.\(^4\) Intellectual has been probed by experimental psychologists who have developed the Stanford-Binet intelligence tests,\(^5\) and disturbances in personality have been studied with the Rorschach\(^6\) and the Minnesota Multiphasic Personality Inventory.\(^6\) Contemporary scientists in nutrition research have explored growth factors, Rose\(^7,8\) with amino acids and McCollum\(^8\) with vitamins—and on and on.

Let me establish my bias in a more forthright manner. To me, the emphasis in osteopathic clinical research concerns investigation of the information from palpatory diagnosis of the neuromusculoskeletal system, and addresses questions stimulated by the clinical experience of body response to manipulative therapy. This single statement takes on a very limited perspective only if you fail to appreciate that it speaks to the interaction of all body systems and their integrative mechanisms in both health and disease as reflected in the somatic component. Palpation provides a major clinical window on this scene.

Why a lack of scientific interest on this scene? First of all, has there been a lack of scientific interest in the somatic system? Certainly not. Even back in the early forties, Hellebrandt\(^9\) reviewed major studies of postural reflexes, biodynamics, and relationships of body mechanics to health. Trauma studies of the somatic structure as a linkage system\(^10\) span the last century, while in the last decade, even in his 1973 Nobel address, Tinbergen\(^11\) drew the attention of the scientific community to considerations of posture with an anecdotal reference to Alexander therapy and its influence on postural configuration. No, there hasn't been any lack of interest in the somatic system, and there hasn't been any lack of recognition of observations of gross asymmetry in posture and movement as initial indicators of structural problems. There has been, however, a failure to acknowledge the use of palpation in a logical diagnostic process for the somatic system, especially when that process has been used to evaluate a diagnostic entity called an osteopathic lesion. Has the failure been one of communication? Yes, I believe it has; especially about osteopathic diagnosis.

A major step to improve communication was taken in the late sixties when the term somatic dysfunction was introduced for use with federal and state agencies and third party insurance carriers, hardly the world of science but still a step in the right direction. Where the term “osteopathic lesion” had communicated little, the term “somatic dysfunction”\(^12\) now appears to have acceptance as a general descriptor for regional disturbance in the somatic system. The two terms, osteopathic lesion and somatic dysfunction, are not necessarily synonymous, but certainly the term somatic dysfunction has begun to communicate the clinical fact of a somatic problem, establishing a focus for its diagnostic documentation and for evidence of its clinical incidence.

Polanyi\(^13\) has suggested that even the most stubborn facts will be set aside if there is no place for them in the established framework of science. The osteopathic lesion has been one of those stubborn facts, and if terminology was a major reason it had difficulty finding a place in the established framework of science, and it did, then a second reason was the lack of documentation of clinical tests of somatic function. Why has this been so crucial to scientific communication? Feinstein’s reference\(^8\) to the physician’s office as a place of scientific investigation is predicated on a careful factual de-
scription of the physical examination tests used to elicit a clinical finding. Early osteopathic literature was more conceptual than factual; physicians elaborated on concepts of what they were finding, rather than on how they were eliciting findings. From *The tyranny of words*, Chase thought probably agree that the concepts reported moved beyond the range of the actual clinical experience; what failed to be reported was a descriptive record of the clinical test performed and a clinical finding elicited. This kind of descriptive research is absolutely essential if a foundation for a clinical science of osteopathic medicine is to be established.

Historically, our own profession's lack of description of diagnostic tests began to be remedied with the writings of Schwab in the 1930s, Allen and Stinson and Beckwith in the 1940s, Beal in the 1950s, and the work of Denslow throughout this entire period. It was Denslow in the 1940s who first applied the knowledge and experience of a clinical test as a bridge to fundamental osteopathic research. He began, on a clinical note, an initial investigation with Korr of spinal reflex thresholds; the Kirksville studies continued successfully far beyond the point of that initial clinical stimulus. This marked the beginning of a new era. During the early period of the profession's development, it had been the osteopathic clinical anatomist who developed the orientation for teaching structure as the basis for osteopathic principles and practice, Halladay, Strachan, and Cathie, to name a few. When Korr presented the concept of a neural basis of the osteopathic lesion, however, the clinician was provided with a physiologic orientation for thinking about palpable findings in the somatic structure, findings that had previously been given largely a positional or anatomic orientation. Now a basis was emerging for consideration of spinal cord segments and mobile vertebral segments, and their clinical reflex relations.

Bowles took immediate advantage of this new knowledge and applied it in his development of a functional model for the clinical testing of patients in office practice. Instead of tests for joints and muscles and range of motion, he perceived passive gross motion tests—gross sidebending, rotation, forward and backward bending—as a means of reproducing the body's everyday movement to serve the clinician as a test pattern for somatic function and dysfunction. Using palpation, the physician could control the demand for motion and evaluate not only the performance of mobile regions, but also the behavior of mobile segmental parts. Experimental scientists may prefer to call this "stressing the system to measure the response." They may even react favorably to the concept of a linkage system that is dynamic, but this requires a new definition, first of the system, and second of the functional links.

Bowles' conception of that system is concerned with patterns of body movement that are already well organized in the central nervous system, with peripheral mobile segments to participate in carrying out the overall commands. Within the framework of this segmental participation there is the expectation that each mobile part be in the right place at the right time to accomplish the goal of response to a command for movement with, of course, minimum energy expenditure and minimum sensory distress. The dynamics are provided by a system of sensory proprioceptive feedback at each segmental level. Again, let me repeat for emphasis, this is not a model that is primarily concerned with the concept of a joint, or an individual muscle. The mobile segment is conceived as a bone with articular surfaces for movement, and adnexal tissues that (1) move it, (2) allow it to be moved from one position to another, and (3) stabilize it in one position. It has a segmental control system that depends upon proprioceptive inputs. This briefly describes an elementary basis in anatomy and physiology for the palpatory examination of a mobile somatic system. In clinical testing the physician examines the system at rest, then initiates and directs the passive introduction of movement to monitor by palpation either the performance of a region or the behavior of a segmental part, depending on test selection.

Findings from palpatory tests that monitor responses to motion have made a sizable contribution to descriptive research about somatic dysfunction. The use of gross movement in test patterns has promoted one very positive basis for communication: descriptions of test results can convey the visually perceivable actions of examiners as they introduce a direction of gross motion to test regional performance. The palpatory criteria have been clarified: criteria for normal response to motion in a mobile spinal segment include compliance and symmetry; for dysfunction, the criteria are resistance (disturbed sense of compliance) and motion asymmetry. A simple comparison of regional response to movement in opposing directions establishes the presence of functional asymmetry as a major sign of somatic dysfunction, and identifies the region. To examine the behavior of segmental parts, the examiner utilizes the same kind of test, inducing gross movement and monitoring the response of motion within one segment before proceeding to the next.

Motion testing criteria have been described also...
Manipulative Techniques
Osteopathic Lesion

PALPABLE FINDINGS:
Structural Irregularity
Tissue Texture Abnormality
Motion Restriction

Fig. 1. Osteopathic terminology, 1930s.

DIAGNOSTIC TESTS

SIGNS: Asymmetry of Structure
Tissue
Motion

Fig. 2. A new basis for clinical communication and recordkeeping, 1960s.

SEGMENTAL DYSFUNCTION

TESTS

1. Initial Screen (IS)
2. Local Scan (LS)
3. Segmental Definition (SD)

SIGNS: Asymmetry - Structure - Tissue - Motion

For segmental dysfunction\(^2\) in a fundamental unit that consists of three segmental components. The clinically significant finding is a central segment that has motion asymmetry; two adjacent segments, above and below, have motion asymmetry that is a mirror image of the finding in the central segment.

With this capsule of descriptive research reviewed, let me go back now to the 1930s (Fig. 1) to indicate our earlier situation, and then outline the development of somatic system diagnosis since that period. The procedures of manipulation were often quite well described. The osteopathic lesion was a local palpable finding in the somatic system characterized by structural irregularities, tissue texture abnormalities, and motion restrictions, and often in the clinical situation by subjective pain and sensitivity to pressure. The space in the center of Figure 1 is not because palpatory tests in osteopathic diagnosis were not used; it is because, up to that time, they had never been carefully described. Allopathic medical diagnosis of somatic pain often consisted primarily of differential diagnosis, which was essentially the ruling out of all other serious local or systemic disease rather than the positive ruling in of a clearly defined somatic problem.

By the late 1960s (Fig. 2), we had the general term somatic dysfunction. Members of a Committee on Uniform Records of the American Academy of Osteopathy, meeting with Kelso as consultant, were quick to recognize the diagnostic signs of asymmetries of structural position, tissue, and motion as basic to the diagnosis of somatic dysfunction. They recommended implementation of a record system\(^2\) using this triad as indicators of problem in the somatic system. By 1970, in spite of the gradual increase in description of isolated tests, there was still a relative lack of organization for a logical sequence of testing procedures to diagnose the problems addressed by manipulation.

In the challenge of curriculum development for the teaching palpatory and manipulative skills at MSU/COM, we established a firm priority for an orderly content of visual and palpatory tests for examination of the somatic system. The goal was to be able to document the location and the characteristics of segmental dysfunction and develop a logical basis for the application of principles for appropriate use of osteopathic manipulation in the treatment program. The model (Fig. 3) follows the general pattern of the regional physical examination: level 1 tests, applicable to each body region, initially screen for signs of systemic problems; they address the question, "does any region of the body show signs indicating problem in a specific
When positive signs indicate problem in the *somatic* system, level 2 tests then scan more carefully throughout the problem region, segment to segment, to determine the specific location of segmental dysfunction. Level 3 tests define the disturbances in tissue and motion that characterize the particular dysfunction identified.

At level 1, the term “initial screen” is widely accepted to describe history and physical examination tests that quickly survey all regions of the body for initial signs of systemic problem. At level 2, the term “scan” is not as widely used in this area of medicine, but the principle of scanning, involving point-by-point observation, seems particularly appropriate for special tests that address this level of differential diagnosis. The initial screen distinguishes the generalized phenomenon of disturbed performance of a somatic region; the local scan elicits the disturbed behavior of one or more individual segments. The probability of their clinical interplay is obvious. The clinical testing procedures to elicit the signs of each, however, are singularly different. Level 3 tests for segmental definition provide the diagnostic detail that is essential for appropriate use of manipulative treatment.

During this same period at COM, Dinnar, a visiting scholar from Israel, showed considerable interest in osteopathic procedures in diagnosis. His focus was on the osteopathic physician as uniquely different from other physicians. Dinnar joined McConnell in asking an important question: What tests do osteopathic physicians perform in diagnosis of the musculoskeletal system that lead to a unique impression of the patient’s problem and a logical use of manipulative therapy in approaching that problem? Together, they looked at the physician and what he does. Analysis of fifteen video-taped actual doctor/patient encounters resulted in a classification of diagnostic tests based on the underlying principle of each test (Fig. 4). In general, the test principles involved in the first three classes are applicable during the initial screen. Class four and five tests are used to scan for segmental location and to define the character of the segmental asymmetry. Class three may be considered somewhat transitional in these respects, depending on the use of the information derived. Dinnar’s conclusion was that tests in classes four and five, especially, were distinctive to osteopathic procedures of diagnosis.

**Comparison of Figure 1 with Figure 5 emphasizes again the basic palpatory signs of osteopathic lesioning and our relative lack of clinical communication regarding palpatory diagnostic tests 50 years ago. The signs of tissue irregularity were certainly well documented. Burns and Cole’s research have described the microscopic features of simulated tissue events. Step-by-step, the diagnostic process has been delineated: there has been a clarification of the triad of diagnostic signs for somatic dysfunction in relation to recordkeeping; in college curricula, we now have the opportunity for a logical sequence of testing procedures for arriving at a diagnosis, not only of somatic dysfunction, but also of segmental dysfunction; Dinnar and McConnell’s classification of the diagnostic tests now includes documentation of fifty tests with samples from each class. These steps in defining the diagnostic process and describing the clinical procedures have been essential developments for basic and applied research. Concurrent with these steps, we proceeded with interexaminer reliability studies of tests from each of the three diagnostic levels outlined, initial screen, local scan, and segmental definition.

Historically, interest in the criteria used by multiple examiners making subjective decisions may precede the development of the Binet-Simon intelligence scales (1905-1913), but this was certainly a good example of the principle in use. Educators developed tests and then individually examined the intelligence of a wide range of student subjects to determine agreement between examiners’ tests scores and the student’s intelligence level as reflected in known performance. This was followed by reassessment of the educators’ tests and their criteria, resulting in the eventual development of the Stanford-Binet intelligence scale as a useful evaluation system for reproducible subjective judgments on intelligence. Our studies to establish the nature of somatic dysfunction have taken a similar design. Three or more examiners have been used, and their judgments, based upon specified tests and defined criteria, have resulted in agreement on aspects of what somatic dysfunction is and where it exists.

Especially in the past 20 years, there has been a growing recognition of the need for testing reliability of medical perceptions. Clinical judgment has been studied in a number of aspects of medical practice: examples are auscultation, electrocardiography, radiology, and ophthalmoscopy. McConnell states the issue clearly when he says that these judgments constitute a very complex, highly subjective measurement that clinical investigators are attempting to convert into the hard currency of reproducible scientific
findings.” In an emerging clinical science such as osteopathic medicine, establishing examiner reliability in palpatory testing procedures plays a fundamental role in providing the base needed for both clinical research and improved clinical care.

The primary goal of interexaminer reliability studies is to establish agreement using a fixed protocol achieving significant agreement in these studies establishes one method in which the tests and the criteria used become valuable, because the findings that emerge are no longer suspect. Simultaneously, findings have an existence that has been carefully defined. These steps have been critical in descriptive communication and have replaced poorly established criteria and procedures for describing palpatory findings. Previously, findings had little credibility, were in fact considered highly subjective and lacking reproducibility. “Moonbeams” was one of the labels that came out of the New York conferences in the 1950s and, although the remark lacked understanding, it may not have been all bias. At that time, the osteopathic lesion was up and standing, but without visible support to the observer lacking experience in palpation.

The purpose of our research has been to define findings used to identify somatic dysfunction. Our interexaminer studies pursued the following general sequence of events. A clinical question provided initial focus for exploring a variety of palpatory tests, with several being selected for study. A training program prepared examiners for testing subjects, who were usually volunteers from both clinical and nonclinical settings. Examination of resulting data raised questions that would form the basis for new directions of interexaminer studies. When statistical analysis provided a level of confidence in test findings of somatic dysfunction, then instrumental measurements were considered for the purpose of quantifying the disturbed function.

Interexaminer reliability studies approach the problem of a subjective judgment by applying the framework and principles of scientific investigation in the following manner:

1. Research designs can be developed in a clinical setting to address a hypothesis according to scientific standards of examiner blind, data retrieval, and statistical analysis.

2. The method of clinical test procedures can be described so that clinical experiments can be repeated by others. The training program for examiners of even one study addresses a significant aspect of this premise. The multiple examiner study is a series of experiments, repeated independently by each examiner. This is the basis for establishing reliability of findings.

3. Criteria for a finding can be clearly described for a subjective judgment—whether that judgment involves subjective interpretation of an x-ray picture, an electrocardiogram record, or a palpable finding on physical examination.

4. When these judgments are made as comparisons of mutually exclusive events (yes/no decisions), then in statistical analysis, the probability distribution for the repeated testing by multiple examiners is described by the binomial distribution (Bernoulli) under assumptions of appropri-
ate sample size, independence of examiners, and random sampling. When sample size is relatively large, this distribution is approximated by the normal frequency distribution, therefore allowing the use of normal statistics.

It is important to recognize in all research that the scope of a finding is immediately limited by the test procedures and criteria that are chosen. Once the method is established, the researcher has placed an experienced and known limitation on the kind of finding to be made—a direction is established—and it is no more limiting in clinical testing than it is in laboratory method.

With attention to these aspects of design, method, and analysis, our early clinical investigations in palpation studied the item of examiner variability. Initially at least, studying the examiner in the use of palpation procedures is like studying an instrument in preparation for its use as a reliable measure of a property to be investigated. Obviously, the expectations of precision for subjective judgment and instrumental measurement are different. What are the factors that influence expectation for reproducibility of findings from palpation tests? Once establishing a focus for what is going to be looked at, and how to look, then an appropriate basis for subjective decision on a simple binomial question can be established; samples are present/absent, comparing right vs. left in testing response to motion, or comparing one segment with an adjacent segment in response to a standardized test. Training programs establish criteria for the kinds of palpable findings to be made and clarify details of the palpation testing procedures to be used. The training program is designed to minimize the possibility that disagreement may be based on procedural errors in the technique of testing, and in recording. Although results of our early studies showed statistically significant examiner agreement on findings from tests for segmental definition that require an advanced palpating tool, they did not reflect the high levels of examiner agreement we might have expected.

It was during this initial experience, however, that we began to recognize an element of variability in the examined subjects that contributed to disagreement among examiners and would affect expectations for agreement. We had asked examiners for decisions on presence/absence of three clinical findings, with each decision based on several tests and multiple palpation cues. Reexamination of some patients indicated that disagreement sometimes resulted from findings that were slight, with one or more motion cues that were transient and varied across examiners. The human subject is not unchanging like the x-ray image or the electrocardiographic tracing. A major variable in clinical tests for response to motion is this transient or slight palpation cue of varying intensity during repeated testing by even a single examiner; this kind of palpation cue continues to vary during subsequent examinations. Our concern was with the identification of a stable quality of palpation cues indicative of clinically significant findings of disease process. At this point, with a newly trained examiner team, we initiated a 3-year study with the same protocol of tests for segmental definition at identified spinal segments. Instead of reporting only three decisions by each examiner, however, examiners reported presence/absence of palpation cues of asymmetric resistance to each motion test at each spinal segment tested, and the direction of the resistance—twenty-seven data points. The first report of this study indicated that in subjects with high agreement levels there was statistical significance to the association between these stable findings of segmental dysfunction and hypertension.

Once confidence was gained in the training process and significant levels of trained examiner reliability were established, we also began to investigate this item of variability in the subject using passive gross motion tests from the initial screen. In a study of six different motion tests, our results revealed that when palpation cues were unstable in a single examiner's repetition of the same test, this variability of transient cues contributed to over 25 percent of the disagreements between examiners on findings of each of the six screening tests. In some tests the transient or unstable findings contributed more than 50 percent of the examiner disagreement. Certainly, subject variability is an important item for consideration in all physical examination tests, and palpation tests of motion are no exception. Biologic variation in living subjects is a major source of errors in repeated measurement.

These results stimulated development of a research design to control for this variable; our selection process identified subjects with major findings that could be expected to be consistent during multiple examinations. Our hypothesis was that agreement levels would be higher in a subsample of subjects previously identified by two independent examiners as having stable major findings of asymmetry during passively induced gross motion tests.

We selected cervical rotation because we had already achieved a significant level of interexaminer agreement on unselected subjects with this test. How high could that agreement be on a sub-
sample of selected subjects, already identified by two independent examiners as having stable findings? Results supported the expectation that a trained examiner can indeed distinguish the specific finding in a subject when an asymmetry to passive gross motion test has been confirmed present in prior examination by another pair of examiners.

Once the determination of clinical tests with reliable findings of somatic dysfunction in the cervical region had been achieved, we were ready to select instrumental measurements with Vorro and Hubbard to quantify aspects of the disturbed function. For this research, motion range and muscle function were studied in relatively asymptomatic subjects. The subjects were selected by agreement of three examiners regarding presence or absence of symmetry in response to a single screening test for cervical sidebending. To address one frequent clinical premise, we would look at evidence regarding contralateral muscular contraction in subjects who were asymmetric, for example, a sidebending right movement limited by contraction of left-sided spinal muscles. Synchronized EMG and videotaped movements were used to gather time, position, and myoelectric data for cervical muscle function during active and passive motion. Three categories of motion were tested: forward and backward bending, sidebending, and rotation. Average age was the same in each of three groups of subjects studied, one group with symmetry, one group with limited response to sidebending right, and one limited to sidebending left. With surface electrodes, Vorro monitored Trapezius, Splenius capitis, Sternocleidomastoid, and three spinal levels of cervicothoracic paravertebral muscles. Results indicated that symmetric and asymmetric groups are indeed different. First, from the kinematic data, comparisons of motion ranges of both asymmetric groups (together) with the symmetric group revealed that all three categories of the cervical motion were limited in the asymmetric groups. From the myoelectric data, there are three parameters that distinguish between the asymmetric groups (together) and the symmetric group. In the symmetric group, muscles began their activity sooner, they reached peak electrical output sooner, and their total electrical output was greater than the asymmetric group. These myoelectric data revealed little to support the specific concept of contralateral muscular restrictors in subjects with left or right asymmetry. The project has provided kinematic and myoelectric measurements that clearly distinguish patients having asymmetry in response to cervical sidebending as a clinical sign of somatic dysfunction. This research continues with frequency analysis of symmetric and asymmetric cervical muscle function and a larger population sample.

The regional motion testing procedures just discussed are used in the initial screen of a patient and only address the question "is there a problem of somatic dysfunction," and they begin to narrow attention to a region. They do not define location of lesioned segments in either the spine or extremity. This is accomplished by tests that scan a region from segment to segment for location of palpable asymmetry of bony position, segmental tissue tension, or segmental response to motion. We selected a pair of palpatory tests that make use of the presence of local deep segmental muscular tension and the associated reduction in segmental mobility. They had proved suitable for identifying segmental dysfunction and for grading its degree of intensity on a scale of 1 least to 3 greatest. A multiple examiner study using this test in the thoracic spinal region was reported at this conference, in March 1981. Statistical analysis of results indicated rejection of the hypothesis that the levels of agreement reached on findings at major and even moderately dysfunctional spinal segments could occur by chance. This study identified a pair of tests for deep segmental tension and reduced segmental mobility that were reliable and had clear-cut criteria for findings that can be used to locate segmental dysfunction.

When lesioned segments can be confidently identified in this way, with predictable anatomic and physiologic parameters, then the related spinal controlled activity can be investigated. Comparisons can be made of tissue properties within and without the lesioned area. Kelso selected thermography as a system with appropriate capability for one of these measurements, namely thermal indicators of cutaneous blood flow. Is blood flow modified in an area of disturbed spinal reflexes, either somatosomatic or viscerosomatic? Are there significant changes in skin temperature when measured before and after manipulative therapy and what does this contribute to knowledge about reflex control of blood flow? Kelso reported on a case study that begins to address these questions at this conference. I believe this is the first occasion reported of instrumental measurement supporting both osteopathic diagnosis and the positive physiologic response resulting from a kind of manipulation that addresses diagnosed segmental dysfunction.

These three measurements, electromyographic, kinematic, and thermographic, are examples of osteopathic applied research in the somatic system made possible by documenting palpatory diagnosis.
of somatic dysfunction and segmental dysfunction.

In practice, experienced osteopathic physicians will often gain initial impressions very quickly through use of screen and scan tests, establish the region and specific location of a major lesioned segment, and then be testing for segmental characteristics as they initiate manipulative treatment. On the other hand, sometimes tests of the initial screen and scan are a stopping point in the diagnosis of dysfunction in the somatic system. If only the location of an area of reduced mobility is established by the clinician, manipulation may never be initiated or may be limited to a more generalized kind of treatment procedure for restoring general mobility, whereas the use of further tests from level 3 would provide increased segmental definition and an opportunity for more specific manipulation, often requiring less operator force.

Level 3 palpatory tests, which go beyond screening and area localization, provide other opportunities for the clinical scientist, and a point of research interest for the basic scientist as well. Larson has carefully described tests for defining soft tissue findings useful in differential diagnosis. He reports how distinguishing the precise location of these tissue changes, and their palpable qualities, may serve the clinician in detecting disturbances in spinal reflex patterns associated with visceral disease. Tests that define segmental motion characteristics offer the same kind of opportunity for differentiating those features of asymmetric lesioned behavior that are frequent (commonplace) from those that are uncommon or unusual. The recognition of an unusual pattern of segmental motion dysfunctions has been the clinical basis for interexaminer studies of the hypertensive patient. This particular pattern of disturbances at specific vertebral and costal segments in the cervicothoracic region exhibits remarkably precise characteristics of asymmetric segmental response to selected motion tests.

Because its distribution follows none of the patterns of curves and crossovers that are more typical of somatosomatic reflex adaptations, and because of its clinical association in the patient who is hypertensive, it emerged as an example of the unusual.

Following on the initial clinical description of this pattern in 1973 two interexaminer reliability studies have been reported. In the first, involving 102 subjects, three trained examiners independently used a selection of level 3 palpatory tests to identify presence/absence of three characteristic vertebral components of this somatic pattern. Results indicated confidence in the reproducibility of the pattern findings and a continuing significance to the clinical association reported between somatic pattern and hypertension. In our second study of 307 subjects, we used agreement of three newly trained independent examiners as the measure of stability for palpatory cues across three consecutive examinations; each examiner performed 27 tests for segmental response to motion. The premise is similar to our studies of level 1 passive gross motion tests; namely, stable palpatory findings provide a positive sign of disease/dysfunction, this time at the diagnostic level of segmental definition. Distribution of agreements is bimodal (Fig. 6), with the lower range of agreements fitting closely to the range predicted by a random model. Within this lower agreement range of 216 subjects with unstable findings, there are 48 hypertensives (22.2 percent). In the upper range, the number of agreements is too great to be explained on the basis of randomness; within this group of 91 subjects with stable findings, there are 48 hypertensives (52.7 percent). Distribution of agreements for the normotensive subsample is not bimodal.

Paramount here is the recognition of a process of clinical selection. This population sample has been divided into groups, not by a single subjective judgment, rather by agreement of three examiners that stable findings are present; the findings are from tests that have been selected because they discriminate a specific complex of spinal dysfunctions. The bimodal distribution indicates a subsample that is different, a group of subjects we can now investigate for other clinical correlates, like blood pressure, and other parameters that further define the disturbed function. When palpatory tests can be used for this kind of selection process, we have a way to study somatic manifestations of visceral disease, as well as the manifestations in visceral tissues of major patterns of segmental dysfunction. The specification of palpatory tests arises from clinical observation. The resulting direction of basic research will rely on knowledge and experience in a field of inquiry, scholarly consideration, and technical study. You may recall my statement at the beginning that these are the qualities for which the clinician reaches out to the basic scientist. In the continuing pursuit of knowledge about authentic somatic relationships, multiple examiner studies contribute a setting in which reliability of clinical findings can be assessed. They help to span the gap between clinical findings and instrumental measurement, between the clinical science and the basic science. In this particular instance, further investigation of reflex relationships is being carried out in thermography studies of the thoracic spinal region comparing patients in renal
failure with normotensive and hypertensive control subjects in a joint project at the Chicago College of Osteopathic Medicine.

I began by saying that palpation provides a window on the scene of somatic function and somatic system interaction with other body systems. In palpation, the clinician often observes some aspect of body condition or of body response before he has any immediate framework of logic to encompass this clinical finding in nature. Especially if stability and predictability become qualities of the phenomenon, he makes a personal commitment to the finding. The basic scientist, on the other hand, may have a model based on reason as foundation for his commitment to a direction of research, before he knows whether it corresponds sufficiently with nature. Whatever the approach, that of the clinical or the basic scientist, the observing becomes scientific only within the context of a clear description of the finding and how it was observed. For the clinician, this is often the least recognized aspect of the clinical science, but still a sine qua non. The future course of osteopathic research will be strongly influenced by clinicians’ ability to provide a logical diagnostic process, a careful description of specific tests, and well-defined criteria for clinical findings.

My goal has been to draw your attention to interexaminer reliability studies of palpation and their use as a tool for clinical investigation. Three levels of a diagnostic process for somatic function have been outlined to assist clinicians in osteopathic descriptive research. In the interexaminer studies reported, tests from the initial screen, local scan, and segmental definition have been singled out, to investigate stable findings of somatic dysfunction and segmental dysfunction as indicators of health problems. For these kinds of studies, a basis for objectivity in a clinical research setting has been set forth. From each test investigated, reliability of palpatory findings has stimulated further directions for applied and fundamental research. Controlled clinical studies of manipulation are now a legitimate prospect, where previously
we have been ill-prepared to take even the first step, that is, diagnostic definition. The fact that some studies of manipulation have already been conducted without this essential first step detracts from their value. To me, the osteopathic profession is on the threshold of another very exciting period of research development. My appreciation is expressed to the Bureau of Research for the opportunity to summarize some clinical observations that will bear on future directions for that research.

I am indebted to Albert F. Kelso, Ph.D., and Myron C. Beal, D.O., for comments and criticism.

24. Kelso, A.F.: Records to assist osteopathic physicians. JAOA 74:751-5, Apr 75
27. Dinnar, U., et al.: Description of fifty diagnostic tests used with osteopathic manipulation. JAOA 81:314-21, Jan 82
33. McConnell, D.G.: Personal communication
42. Johnston, W.L., et al.: Multiple examiner study of spinal segmental dysfunction. JAOA 80:752-3, Jul 81
43. Johnston, W.L., et al.: Multiple examiner study of spinal segmental dysfunction. JAOA 80:752-3, Jul 81
44. Johnston, W.L., et al.: Multiple examiner study of spinal segmental dysfunction. JAOA 80:752-3, Jul 81
47. Johnston, W.L., et al.: Multiple examiner study of spinal segmental dysfunction. JAOA 80:752-3, Jul 81
52. Johnston, W.L., et al.: Multiple examiner study of spinal segmental dysfunction. JAOA 80:752-3, Jul 81