REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	28.Sep.00	DISSERTATION

4. TITLE AND SUBTITLE
CERVICAL RADICULOPATHY AND CARPAL TUNNEL SYNDROME: A PROSPECTIVE DETERMINATION OF THE RELIABILITY, DIAGNOSTIC ACCURACY, AND PREDICTIVE VALIDITY OF COMMONLY USED CLINICAL

5. FUNDING NUMBERS

6. AUTHOR(S)
MAJ WAINNER ROBERT S

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF PITTSBURGH

8. PERFORMING ORGANIZATION REPORT NUMBER
CY00379

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS	15. NUMBER OF PAGES 227
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

Dear AFIT/Program Manager (Maj. Russo)

I, Maj. Robert S. Wainner, BSC, undertook and completed Doctoral studies at the University of Pittsburgh from 8/97 to 8/00, and request reimbursement for my dissertation ($200.00), a copy of which has been forwarded along with this request. The check can be sent to my home mailing address below:

Maj. Robert S. Wainner
231 Oak Leaf Dr.
San Antonio, TX 78209

Hm Ph. (210) 821-6620

CERVICAL RADICULOPATHY AND CARPAL TUNNEL SYNDROME: A PROSPECTIVE DETERMINATION OF THE RELIABILITY, DIAGNOSTIC ACCURACY, AND PREDICTIVE VALIDITY OF COMMONLY USED CLINICAL EXAMINIATION MEASURES AND PATIENT SELF-REPORT INSTRUMENTS

By

Robert S. Wainner

B.S. School of Allied Health Sciences, University of Texas Medical Branch, 1985

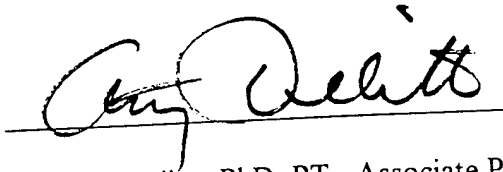
M.S. School of Allied Health Professions, University of Kentucky, 1992

Submitted to the Graduate Faculty of the School of Health and Rehabilitation Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy

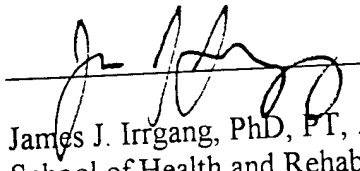
University of Pittsburgh

2000

Committee Signature Page



Anthony Delitto PhD, PT, Associate Professor and Chair, Department of Physical Therapy, School of Health and Rehabilitation Sciences, University of Pittsburgh.



James J. Irrgang, PhD, PT, ATC, Assistant Professor, Department of Physical Therapy, School of Health and Rehabilitation Sciences, University of Pittsburgh.

Julie M. Fritz, PhD, PT, ATC, Assistant Professor, Department of Physical Therapy, School of Health and Rehabilitation Sciences, University of Pittsburgh.

Michael L. Boninger, MD, Associate Professor and Research Director, Department of Physical Medicine and Rehabilitation, University of Pittsburgh Medical Center. Medical Director of the Human Engineering Research Laboratories

CERVICAL RADICULOPATHY AND CARPAL TUNNEL SYNDROME: A PROSPECTIVE DETERMINATION OF THE RELIABILITY, DIAGNOSTIC ACCURACY, AND PREDICTIVE VALIDITY OF COMMONLY USED CLINICAL EXAMINIATION MEASURES AND PATIENT SELF-REPORT INSTRUMENTS

Robert S. Wainner, PhD

University of Pittsburgh, 2000

Background: Patients with cervical radiculopathy and carpal tunnel syndrome result in significant medical and occupational costs annually. There is a need to establish cost-effective, reliable, and accurate means for the diagnosis of both conditions. The purpose of this study was to determine the reliability, diagnostic accuracy, and predictive validity of items of the clinical examination used for the diagnosis of cervical radiculopathy and carpal tunnel syndrome. **Methods:** Forty-one females (mean age 44 ± 12.5 years) and 40 males (mean age= 45.0 ± 11.4 yrs) received a standardized electrophysiological examination of their affected limb. Patients received a diagnosis based on their presenting symptoms and electrophysiological examination. Two physical therapist raters, blinded to the results of the previous exam and suspected condition, performed a standardized clinical examination of the same limb. At six-weeks, all subjects were mailed the same self-report forms initially completed at the time of enrollment. **Results**: Thirteen subjects (16%) and 31 (38%) subjects were diagnosed with cervical radiculopathy and carpal tunnel syndrome, respectively. The following levels of reliability were found for 54 different clinical examination variables assessed in this study: 9 were Excellent (Kappa >.75), 33 were Fair to Good (Kappa= .40 - .75), and 11were poor (Kappa <.40, ICC <.75). Twelve test for cervical radiculopathy and six tests for carpal tunnel syndrome had acceptable Likelihood ratios. Only question number 7 for cervical radiculopathy had a definitively acceptable Likelihood ratio (6.5, 95CI= 2.3 - 18.0). Seventeen surgical predictors had acceptable Likelihood ratios. Based on patient global rating of change, 12 predictors of worsened status and 18 predictors of improved status for non-surgically treated carpal tunnel subjects had acceptable Likelihood ratios. For each diagnostic and predictive condition, a single test-item cluster was identified that would produce post-test probability changes in this sample of subjects that ranged from 23% to 69%. **Conclusion**: The majority of test items in this study had acceptable reliability. None of the definitive LR+ values for individual tests or test clusters had a lower bound that would result in post-test probabilities larger than 33%. More precise estimates are required to establish the diagnostic and predictive validity of clinical examination tests for cervical radiculopathy and carpal tunnel syndrome

Thank you, Jesus Christ, for life, salvation, and the ability You so graciously granted me to complete my Doctoral program.

Thank you, family. You have been supportive and patient beyond compare. Diane, you are a treasure: I only hope I can inspire and encourage you as much as you have inspired and encouraged me. Rachael, Rebecca, Andrew, Bethany, and Joshua: thank you for the light-heartedness, perspective, and frequent "reality checks" you brought and continue to bring into my life. Jordan Lee, thank you teaching me not to take myself too seriously, for teaching me how to be a better "juggler", and for what you have taught all of us about the true meaning of life in such a short period of time; your family welcomes you! Mom and Dad, Coach and Mrs. Bonner, I could not have come this far without you and am indebted to you all.

Thank you, Committee Members, for your guidance, sacrifice, and support. Julie, a special thanks for providing me with a role model of what a Doctoral student should be as well as your close mentorship and friendship.

Thank you, to the many other people who have made my academic accomplishments here at the University of Pittsburgh possible. Please grant me your forgiveness as space does not allow me to properly acknowledge your invaluable contributions.

TABLE OF CONTENTS

List Of Tables

List of Figures

1.0 INTRODUCTION

This multi-center, prospective, diagnostic-test study proposes to evaluate the efficacy of commonly used clinical examination and patient self-report measures to diagnose and predict outcome in patients with suspected cervical radiculopathy (CR) and suspected carpal tunnel syndrome (CTS). This study will also evaluate the efficacy of these same clinical examination and patient self-report measures to predict patient outcome at 6wks. Individuals with suspected unilateral CR and suspected primarily unilateral CTS will undergo a standardized electrophysiological examination (EMG/NCS) of the affected upper quarter. Following the EMG/NCS examination, patients will also undergo a battery of clinical examination measures and patient self-report measures which will then be repeated by a second examiner. Patient status will be determined at 6wks by a mailed questionnaire that includes a global-rating of change scale and asks the patient if they have received surgical intervention. The specific aims of this research are the following:

1. Inter-rater reliability: clinical examination measures commonly used to evaluate patients with suspected CR or with suspected CTS will demonstrate good (K=.60 - .75, ICC=.75 - .90) to excellent (K= >.75, ICC= >.90) levels of test-retest reliability when the same patient is evaluated by two different physical therapists

2. Test Diagnostic Accuracy: individual items from the clinical examination measures and patient self-report instruments will demonstrate acceptable diagnostic accuracy values (Sn or Sp $\geq$ .70 or LR+ $\geq$ 2.0 or LR- $\leq$ .50) for their respective condition (CR or CTS) when compared to a neural impairment reference criterion.

3. Test Predictive Validity: individual items from the clinical examination measures, patient self-report instruments, and the EMG/NCS findings will demonstrate acceptable diagnostic accuracy values (Sn or Sp $\geq$ .70 or LR+ $\geq$ 2.0 or LR- $\leq$ .50) for their respective condition (CR or CTS) when compared to a patient outcome reference criterion.

4. Test Item Clusters (TIC): an optimum and parsimonious cluster of test items from the clinical examination measures and patient self-report instruments will be identified and

demonstrate acceptable diagnostic accuracy and predictive validity values (Sn or Sp $\geq$ .70 or LR+ $\geq$ 2.0 or LR- $\leq$ .50) for their respective condition (CR or CTS) when compared to a neural impairment reference criterion and when compared to a patient outcome reference criterion.

2.0 STATEMENT OF THE PROBLEM

This section will cover four topics: 1) the impact of CR and CTS on society; 2) reference criteria for CR, CTS, and patient outcome; 3) diagnostic tests considerations; and 4) critical appraisal of existing of clinical diagnostic test technologies for CR and CTS.

2.1 The Impact of Cervical Radiculopathy and Carpal Tunnel Syndrome on Society and on Diagnostic Decision Making

Patients with neck pain and CTS are frequently encountered in primary care,[1] physical therapy,[1,2] and a variety of medical specialty practices that include orthopedics, physiatry, neurology, and neurosurgery. The prevalence of neck pain has been reported to range between 16-18% for the middle aged population and approximately 10% of the population will develop neck pain, with or without referral of pain into the upper extremities, during any given month.[3] Estimates of the number of people who will have at least one episode of neck pain in the course of their lifetimes range from 33[4] to 50%.[1] Although the exact number of patients who develop chronic neck pain is unknown, there is some evidence that it may be substantial. Thirty-two percent of patients in a large, prospective study were noted to have moderate or severe neck symptoms at a minimum 10 year follow-up period.[5] Cervical radiculopathy, which most often occurs as a result of irritation and compression from a herniated cervical disc or osteophyte,[6] is but one of many possible disorders that can give rise to neck pain and disability.[7,8] However, data related to socioeconomic impact of neck pain as a result of CR could not be located. Because CR is thought by some to be one manifestation of neck pain resulting from a degenerative continuum,[7,8] the pain and disability specifically attributable to CR may be considerable.

CTS is the most common nerve compression disorder of the upper extremity with reported prevalence rates ranging from .1 to 2% in the US population[9-11] and affects as many as 2%[9] to 15% of workers in high risk industries.[12] In addition to the frequency of occurrence, CTS treatment complications and the percentage of patients with recurrent symptoms are sobering. Approximately 200,000 patients undergo surgical release of the volar carpal ligament annually.[13] According to Mackinnion's review, 7%-20% of these surgically treated patients fail to obtain relief[14] and the percentage of patients who experience a recurrence of symptoms after steriod injections into the carpal canal ranges from 8%–94%.[15]

Cervical radiculopathy and CTS can produce similar signs and symptoms that make distinguishing between the two conditions difficult.[9pp.10-11] These signs and symptoms may include pain, sensory disturbances, and weakness of the upper extremity.[16,17] In addition, there is evidence that a small percentage of patients with these symptoms are affected by both conditions concomitantly.[18-20] Due to the similarity of presentation in patients with cervical radiculopathy and CTS, many of the same examination measures are often used to evaluate patients suspected to have either condition. This is done in an attempt to differentially diagnose or discriminate between the two and thereby "rule-out" one condition or the other.[16,21,22] However, unless the diagnostic properties of a given test or measure are known, differential diagnosis and informed decision making cannot occur in a quantifiable and interpretable manner.[23,24pp120-125] Unfortunately, the diagnostic properties of tests and measures used for the clinical examination of patients with suspected CR or CTS are largely unknown or not well established.[25]

2.2 Diagnostic Tests Considerations

Advances in technology and the availability of sophisticated laboratory tests have increased our diagnostic power for certain disorders. This selective increase in diagnostic power and reliance on quantitative diagnostic tests have led some clinicians to view data obtained by these procedures as "hard", or objective, and data obtained from the clinical examination as "soft", or subjective.

This viewpoint has led many clinicians to rely on clinical laboratory tests for establishing a diagnosis. However, data should be judged by their power and not by their appearance.[24pp.19-21]

It is clear that both CR and CTS can result in a substantial amount of suffering and disability. In addition, both conditions result in significant medical and occupational costs annually. [9,26pp.10-11] There is a definite need to establish cost-effective, reliable, and accurate means for the diagnosis of both conditions. Aside from accessibility and ecomomic considerations, these tests would be even more valuable if they were useful for predicting patient outcome. The effort required to develop and identify such tests is formidable: appropriate research methodology must be employed; an adequate gold standard to determine presence of condition and patient outcome must be identified; and diagnostic test properties must be reported using metrics that allow for quantification of test results and their probabilistic interpretation. .

The clinical examination, which consists of history, physical examination, and manual test procedures, is once again increasingly relied upon in this era of medical cost-cutting.[27] There are four purposes or activities for which the clinical examination, in particular the history, has been shown to be a an extremely powerful tool. [28-30] These four purposes are: making a diagnosis; ruling out diagnostic hypotheses; identifying disorders in early stages; and establishing a prognosis.[31] Indeed, with the exception of patients suffering from endocrine and alimentary disorders, information obtained from the history alone has been shown by several studies to be sufficient for establishing a diagnosis 63- 88% of the time in patients seen at outpatient medical clinics.[28-30] The physical examination of the patients in these studies provided enough information to establish the diagnosis in most of the remaining cases and routine and laboratory tests contributed to the diagnosis only in 3-14%.[29] The ability of the clinical examination to predict how patients would be managed produced similar results.[28.29] Another example of the diagnostic power of the history is a battery of four specific questions called the CAGE which are related to drinking behaviors. This particular battery of questions is

more sensitive and specific than any laboratory or physical examination finding for the diagnosis of alcoholism.[32]

Despite the demonstrated value of the clinical examination, investigations of the precision and accuracy of the clinical examination have lagged behind similar studies of laboratory tests.[31] Sackett gives five possible reasons for this: 1. Such investigations are challenging to design and arduous to execute; 2. Clinical diagnoses seldom reside in a single finding but rather are usually derived from a pattern or cluster of findings; 3. A lack of interest by clinical investigators in true clinical research; 4. Pecuniary interests in high technology research; and 5. Belief by many physicians that the "art" of diagnosis is incapable of being elucidated and defined by scientific investigation.[31] Recently there has been a renewed emphasis on the clinical examination. The Journal of the American Medical Association now publishes an ongoing series of articles entitled "The Rational Clinical Examination Series" that is devoted to research of the clinical examination.[27] International groups have also been established whose goal is fostering research efforts of clinical examination procedures by providing information and a collaborative forum for clinical investigators.[33]

Very little high-quality research has been reported regarding the diagnostic properties of specific clinical examination procedures for patients with disorders of the neuromuscular skeletal system. Despite the numerous text books devoted to the description and application of diagnostic tests for neuromusculoskeletal lesions,[34,35,35] descriptions of the diagnostic properties of the tests are almost uniformly omitted.[36] However, the lamentations over the current knowledge fund and calls for research ring hollow when there is no plan.[27] This study will assess the reliability, diagnostic accuracy, and predictive validity of several common clinical examination measures and patient self-report instruments used to evaluate patients with suspected CR and suspected CTS

2.2.1 Levels of Efficacy

The primary purpose of diagnostic tests is to provide clinical information which can discriminate among disease states and thereby improve patient management.[37] However, other purposes of diagnostic tests include screening asymptomatic individuals for disease, monitoring the course of a disease, and establishing a prognosis.[37,38] Fineberg has proposed a hierarchical approach to the assessment of diagnostic tests[39] that has been expounded upon by Shwartz[37] and Deyo et. al.[40] This hierarchical approach consists of evaluating diagnostic tests at different levels of efficacy. These levels of efficacy are categorized as technical, diagnostic, therapeutic, and outcome and are described below.
Technical: Refers to the ability of the test procedure to demonstrate adequate safety, be accessible to patients, and have reproducible results.[40] Inter-rater reliability is one measure of a test's technical efficacy and is a pre-requisite for establishing test validity.[41]
Diagnostic- Diagnostic tests are utilized to determine the presence of a target disorder in either asymptomatic or patient target populations. Diagnostic accuracy is usually assessed by comparing a test's results with those of an external reference standard.[38] The external reference or "gold standard" used for comparison is the most accurate and appropriate method of determining the presence or absence of a target disorder and is usually costly and/or involves a moderate to high degree of risk.[38,40] Therefore, clinicians utilize diagnostic tests that are less costly and involve lower risk but are still effective.
Therapeutic & Outcome- These are the highest levels of efficacy for a diagnostic test and arguably the most important. A highly accurate diagnostic test is no guarantee that the test is useful. The true value of a diagnostic test is the ability to determine a course of treatment or predict treatment outcomes through its application.[37,38] Another aspect of outcome efficacy is the cost effectiveness of a diagnostic test in comparison to alternative diagnostic strategies.[40]

The technical and diagnostic levels of efficacy for tests and measures included in this study will be assessed. In addition, follow-up data collected at 6 weeks will allow an approximation of the outcome level of efficacy for the tests and measures assessed in this study. None of the diagnostic tests for CR considered in this study and only a few tests for CTS have been assessed at the therapeutic or outcome level of efficacy.

2.2.2 Research Methodology

The most appropriate research study design for an investigation is determined by the question being asked.[42] For example, the randomized clinical trial is considered the paragon for assessing the effectiveness of a treatment.[43] Similarly, optimum methodological principles have been proposed to assess the efficacy of a diagnostic test.[37,44-46] There are three basic considerations when assessing the diagnostic properties of a test. The first is the gold standard or reference criterion to which the test in question is compared. The second is the spectrum of patients to which the test is administered or applied. The third and final consideration are the procedures used to control bias. Each of these considerations will be discussed below.

Gold Standard- The gold standard serves as a reference criterion by which properties of the diagnostic test in question are determined. Although the gold standard is more accurate than the test being compared to it, is also usually more costly, more time consuming, and involves more risk to the patient;[38] hence the need to develop a simpler and less costly diagnostic test that can accomplish the same purpose with minimal loss of accuracy. Procedures that define anatomic and physiologic abnormalities, including surgery, are often used as gold standards.[40] Other less conventional gold standards include expert clinician opinion and clinical course or outcome.[47,48] All gold standards, no matter how good, have some degree of imperfection[37,45,49] and what constitutes the single "best" gold standard is often the subject of much debate.[40] Resolving these dilemmas may depend on the intended clinical use of the diagnostic test being assessed and the best available standard may often be "silver, bronze, or tin" in hue instead of "gold".[40]

Patient Spectrum- This term refers to the range of features found in the patient sample used to challenge or assess the diagnostic properties of a test.[44] The pathologic, clinical, and co-morbid components of the target disorder must be considered when assembling the patient sample that will be used to assess the diagnostic test being evaluated. The *pathologic* component refers to the extent of disease process, such as localized versus extensive cancer. The *clinical* component refers to features such as chronicity and severity of symptoms. The *co-morbid* component refers to co-existing pathology unrelated to the disease of interest. Each component may adversely affect the positive or

negative diagnostic accuracy of the test in an unpredictable fashion, depending on the disease and diagnostic test in question. For example, a test that performs well with patients whose disease process is mild may perform poorly with patients who's disease process is advanced.[50] Patients who serve as controls should have conditions with pathologic features or similar signs and symptoms that might be easily confused with the disease of interest. Including these types of patients as controls is useful for assessing the number of false positives a test will yield and thus provides a meaningful interpretation of test specificity.[44] Almost any test can distinguish between severely diseased patients and healthy control subjects. The true challenge of test validity occurs when a study includes control subjects that resemble the population of patients to which the diagnostic test will be applied in clinical practice.[45]

Biases- For each patient, the investigator must determine whether the diagnostic test is positive or negative and if the disease condition is present or absent. If these determinations are not independent, a false index of test diagnostic accuracy may result. Control must be exerted for several types of biases that include: work-up, diagnostic review, test-review, and incorporation.[44] Different synonymous descriptors have been used by other authors to describe these biases.[40,45] *Work-up bias* occurs when the result of a test affects the subsequent clinical work-up needed to establish the diagnosis of the target disorder. For example, a patient with a negative test may have a less intense work-up or may not even have the gold standard procedure applied to them since they are thought to be disease free based on the results of the test. This type of bias can lead to under diagnosis but not over diagnosis. *Diagnostic-review bias* occurs when the result of the diagnostic test being assessed affects the determination of whether the target disorder is present or absent and may result in over diagnosis as well as under diagnosis. *Test-review bias* occurs when the presence or absence of the target disorder is known to be established and affects the subjective interpretation of the diagnostic test being assessed and can also lead to over diagnosis or under diagnosis as well. *Incorporation bias* occurs when the test in question is incorporated into the evidence used to establish the presence of the target disorder.[44]

Other potential difficulties and issues to consider when assessing the accuracy of a diagnostic test include: inter-rater reliability; whether the test was performed singly or in combination with other tests; what metrics were used to quantify test efficacy; if the test procedure was operationally defined; and if the setting and population it was applied to were clearly defined.[38,40,44,45]

Although many of the preceding issues seem straightforward and intuitive, it is clear from the literature that sound methodological criteria are often not adhered to when assessing diagnostic tests. Sheps et. al.[51] reviewed 129 articles against 7 methodological criteria identified as being important for diagnostic test research. Overall, 74% of the studies failed to adhere to more than four of the seven criteria and revealed the following: 68% employed a well-defined gold standard; 32% operationally defined how tests were interpreted; interpretation of test results was blinded in 40%; approximately 20% used the terms sensitivity and specificity incorrectly; and the influence of disease prevalence and practice setting were considered in only 19%.[51] A qualitative review of the literature dealing with the accuracy of diagnostic tests for low back pain revealed major methodological shortcomings in most studies and only 19 out of 36 articles scored over 55 out of 100 points.[52] Research methodology employed in the development of diagnostic tests must possess the same rigor currently required for clinical trials of treatment effectiveness. Not adhering to sound methodological criteria may result in improper patient management[44] and a confounding of clinical treatment trials because of an inability to properly define the patient population and assemble a homogeneous patient sample.[27,53]

2.2.3 Metrics and Interpretation of Test Properties

Each component of the clinical examination can be considered a separate diagnostic test.[27] Once the clinical examination is performed, the clinician interprets the findings both individually and collectively in the clinical decision making process. Determining the relevance of the clinical examination findings in a meaningful fashion requires three mechanisms: first, a means of establishing a significant probability or association between an item or items of the clinical examination and the target disorder; second, a means of determining how much the result contributes to the diagnosis above and beyond other clinical examination results; and third, a means of determining if the test results indicate an increased or decreased chance of the target disorder being present, beyond that expected prior to testing.[23.24pp120-125] Three types of metrics used to determine the relevance of the clinical examination findings have been described and will be discussed below.[24pp.69-15254,55]

Sensitivity and Specificity- Test sensitivity (Sn) and specificity (Sp) are conditional probabilities that can be used to define the informational contribution of a test.[44,45] Test Sn is defined as the probability of obtaining a positive test result when the target disorder is present. Likewise, test Sp is defined as the probability of obtaining a negative test result when the target disorder is absent.[24pp.81-82] Sensitivity and Sp calculations are illustrated in Figure 1.

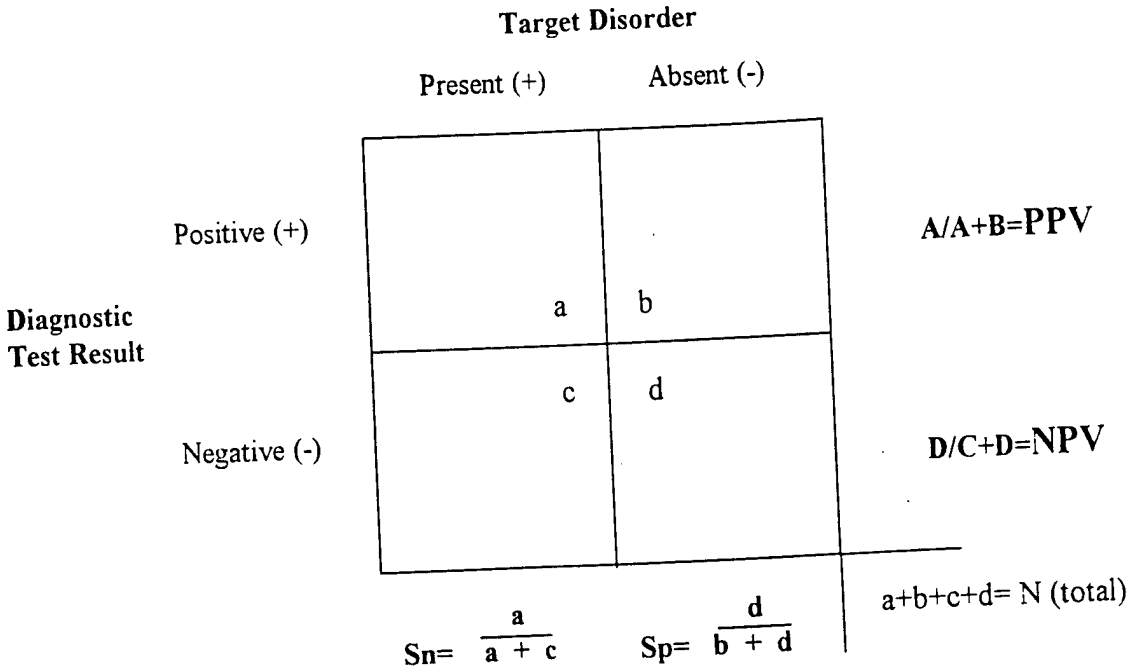


Figure 1 Sensitivity and Specificity

The Sn or Sp of a test depends in part on the intrinsic properties of the test and in part on the threshold criteria used to establish a positive or abnormal test result. Although it is desirable for a test to have both high Sn and Sp, factors that contribute to improving one respective proportion often mitigate the other.[24pp 105] A single test that results in a dichotomy (present/absent, positive/negative) will have only one Sn and Sp value. Tests that produce ordinal or continuos results have many possible Sn and Sp values, depending upon the threshold criteria chosen to define a positive or negative test.[54] Sensitivity and specificity may also be increased or decreased for dichotomous tests by combining the results of two or more dichotomies and treating this cluster as a single diagnostic test. To increase specificity, for example, two out of three tests may be required to be positive in order for the single test cluster to be considered a positive diagnostic test. The same procedure could be used in a similar but opposite fashion to increase test sensitivity by minimizing the requirements for a positive test cluster.[24pp.105.23]

The threshold criteria is established depends on which test property is more desirable (Sn or Sp). Whether Sn or Sp is desired depends on both the intended purpose of the test (to screen or diagnose) and the consequences of intervention.[24pp.99] For example, a test used to screen for cancer should be highly sensitive in order to prevent a case from being missed. Specificity may be sacrificed in order to increase test sensitivity because the consequences of a missed case are disastrous compared to the cost and discomfort of the subsequent work-up for patients who have a false positive test finding. Likewise, a test used for the diagnosis of a target disorder should be highly specific if surgical intervention is based in whole or in part on the result of the test. In this case, some sensitivity will likely be sacrificed in order to increase test specificity because the consequences of a false negative finding may be only minimal when compared to the increased morbidity associated with a false positive test.[38pp.105-108]

Predictive Values- Unfortunately, Sn and Sp can be evaluated only if the true health status of the patient is known. In practice, the clinician rarely knows a-priori if the target condition is present in the patient he or she is evaluating, otherwise the diagnostic test would be unnecessary. Therefore, Sn and Sp are of limited value for determining the probability of whether a patient is more likely or less likely to have the target condition based on the result of the test.[38pp.85] The real question of interest that must be answered is "If a patient has a positive or negative test, how likely is he or she to have the disease?"[44] One method of determining this probability is the calculation of predictive values. The positive predictive value (PPV) measures the pre-test probability that a patient actually has the target disorder when the test is positive. Likewise, the negative predictive value (NPV) measures the pre-test probability that the patient does not have the target disorder when the test is negative. The terms pre-test probability and prevalence often are used interchangeably; the former is used to describe individuals and the latter when describing groups.[45]

Calculation of PPV and NPV is illustrated in Figure 1. Like Sn and Sp, predictive values are of limited clinical use but for a different reason; they are calculated from left to right in the 2 X 2 contingency table and are therefore dependent upon disease prevalence

which makes them unstable. Regardless of the test's Sn or Sp properties, as prevalence falls the PPV must fall along with it and the NPV must rise. Likewise, when prevalence rises so does the PPV while the NPV falls.[24pp.88] The dependence upon disease prevalence and the unstable nature of positive and negative predictive values is illustrated in Figures 2 and 3 on the following pages. In the past, predictive values for an estimated prevalence rate of 50% were often given as a standard characteristic for a test. Because predictive values are prevalence dependent, they are useless in other settings where the prevalence or pretest probability of the disorder is different.[54] Therefore, clinicians must match a patient's history specific prevalence to the Sn/Sp values of a given test in order to then derive clinically meaningful predictive values,[36] which can be quite cumbersome if not impractical in a clinical setting.[54]

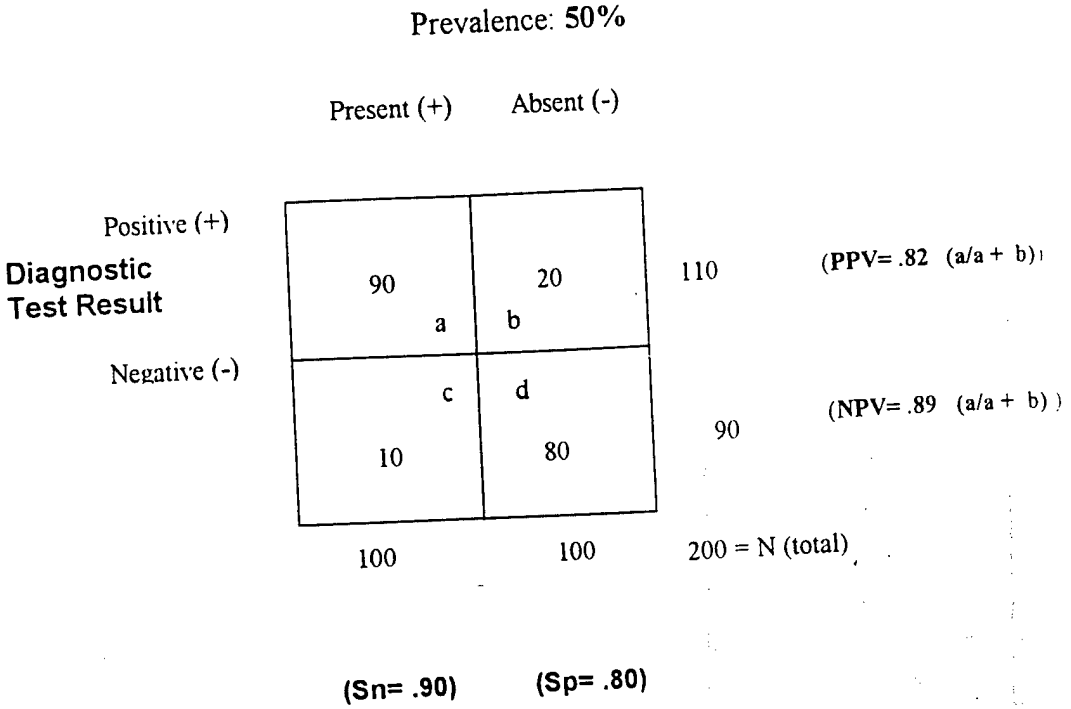


Figure 2 Predicitive Values, 50% Prevalence

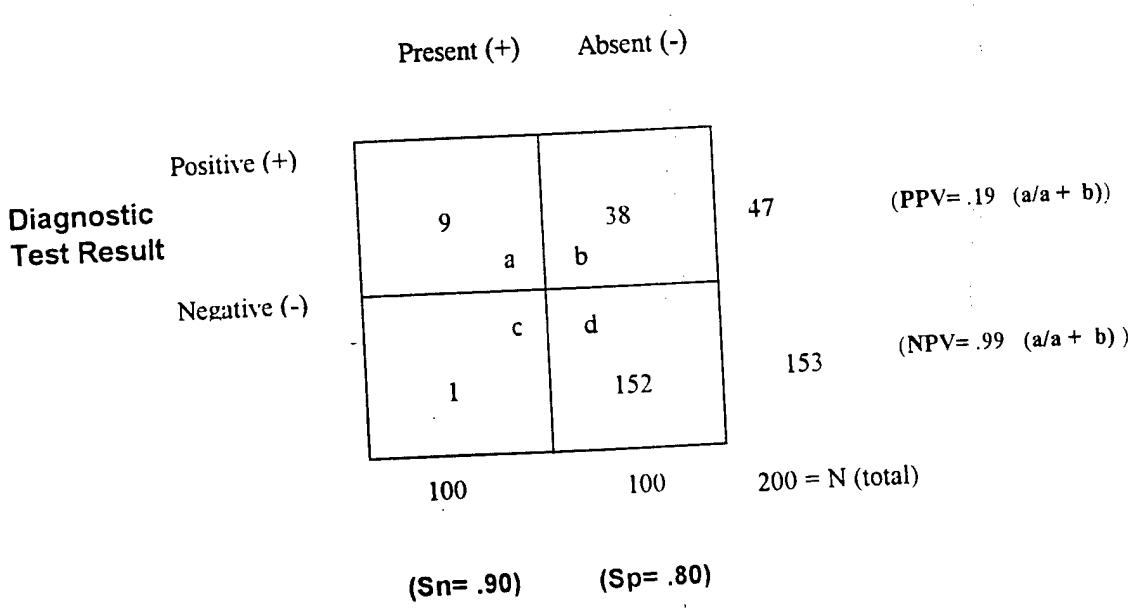


Figure 3 Predicitive Values, 5% Prevalence

Likelihood Ratios- Use of a likelihood ratio (LR) is another method for determining the probability of whether a patient is more likely or less likely to have the target condition based on the result of the diagnostic test. The concept of a LR has been advocated as a better means for assessing the properties of a diagnostic test and as a practical, valuable tool for clinical decision making.[54] An LR is a ratio of two probabilities that expresses the odds that a given level of a diagnostic test result (positive or negative) would be expected in a diseased patient compared with a non-diseased patient[54.38p.120] and is illustrated below.

$$LR = \frac{\text{Probability of test outcome given diseased patients}}{\text{Probability of test outcome given non-diseased patients}}$$

When an LR exceeds 1, the odds favoring a disease increases; when the LR becomes less than 1, the odds favoring the disease decrease; and when an LR approaches 1, the odds favoring a disease do not change and the test is indeterminate.[56] Positive (LR+) and negative (LR-) LR's algebraically combine Sn and Sp to describe more than the independent values themselves;[55] they summarize the information of both Sn and Sp and thereby represent the discriminative power of a test. Positive and negative LR's are computed in the following manner:[54]

$$LR+ = Sn/(1-Sp)$$

$$LR- = (1-Sn)/Sp$$

The following example based on a study by Fritz et. al.[57] is helpful for illustrating the interpretation of LR's:

A treadmill walking test (longer walking time during inclined walking) is used to diagnose patients suspected of having lumbar spinal stenosis. The treadmill test LR+=

6.49 and the LR-= .54. This means that a positive treadmill test is 6.49 times more likely to occur in patients *with* lumbar spinal stenosis than from those *without* lumbar spinal stenosis. Similarly, a negative treadmill test is only .54 times as likely to occur in patients *with* lumbar spinal stenosis than from those *without* lumbar spinal stenosis.

Several authors have described three important properties or advantages of LR's:[24,54,55pp120-123]

1. Likelihood ratios are stable. Because they are calculated vertically in the 2 X 2 contingency table, LR's do not change with changes in the prevalence or pretest probability of the target disorder.
2. Likelihood ratios may be established for multiple levels of test outcome. Establishing multiple level LR's improves their diagnostic properties for test results that are ordinal or continuous scaled.
3. Likelihood ratios allow a clinician to immediately assess the impact of a test result on the posttest probability that a patient will have the disease of interest and can guide sequential testing. If the pretest prevalence (or probability) of a disease is known or can be estimated, the posttest probability of the disease being present can be calculated using the formula below which is be derived from Bayes theorem:

Pretest odds * Likelihood Ratio= **Posttest Odds for the Target Disorder**

Where: Prevalence/(1 – Prevalence)= **Pretest Odds**

Because clinicians may be more comfortable with probabilities than odds, the posttest odds may be converted back to a probability in the following manner:

Posttest Odds/(1 + Posttest Odds)= **Posttest Probability**

Once again, the spinal stenosis example from Fritz et. al. is helpful for demonstrating how the LR of a test (LR+ in this case) may be used to change the probability estimate for the presence or absence of a disorder in a given patient:[24pp.123-126]

Diagnostic test LR+= 6.49 The estimated pretest probablity of the disorder= 40% or .40

Test performed and result is (+)

Convert to pretest odds: **.40/1 - .40= .40/.60= .67**
Pretest odds= .67

The pretest odds for the target disorder	X the LR for the diagnostic test result	=	The posttest odds for the target disorder
= .67	**X 6.49=**		**4.35**

Convert posttest odds back to posttest probablity: **3.35/3.35 + 1= 3.35/4.35= .81**

Post test probability= .81 or 81%

In the example above, the pretest probability of the patient having the target disorder prior to the test result was equal to the estimated prevalence rate of 40%; the positive diagnostic test result has now increased the probability to 81%. If another test is performed, the pretest probability for the target disorder would now be 81%. Provided the tests are independent, this sequence of testing and adjusting the posttest probability may be continued until the clinician is comfortable deciding whether the target disorder is present. For an LR-, the same process can be carried out to adjust the posttest probability of the absence of the target disorder.

Three disadvantages of LR's have also been reported and include the following:[54,55]

1. Knowledge of a test's Sn and Sp is still required. Because the same LR can be the result of the combination of very different Sn and Sp values, the Sn/Sp of a test must be known when false positives or false negatives are to be avoided as much as possible.

2. The posttest probabilities generated by LR values are not linear; the discriminative strength (i.e. resultant posttest probabilities) of an LR+ value of 10 is not ten times that of an LR+ value of 10.

3. The precision of LR's depends on the proportion of diseased and non-diseased subjects. The confidence interval around an LR becomes progressively wider as the imbalance between diseased and non-diseased subjects increases.

Another potential disadvantage is the burden for clinician to establish posttest probabilities. The need to convert back and forth between pretest probability/pretest odds and posttest odds/posttest probability can be confusing and somewhat time consuming. However, this problem is easily remedied with the use of Fagan's nomogram (Figure 4).[58] Once the prevalence of a disorder has been estimated and the LR's of a given test are known, the posttest probability can be determined by using a ruler and the nomogram.

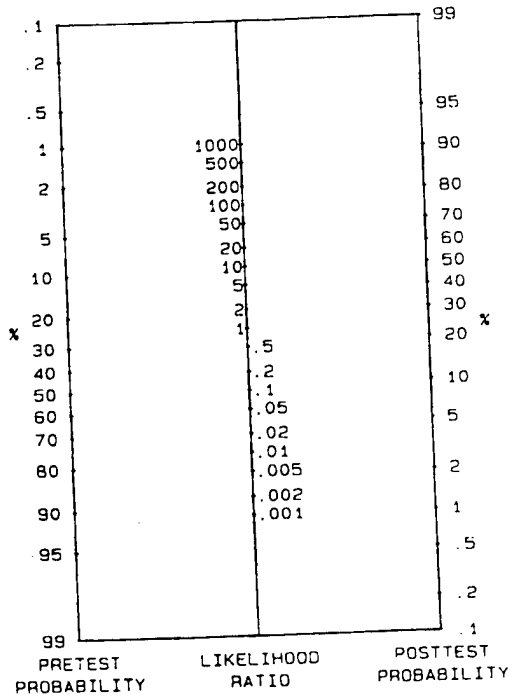


Figure 4 Fagan's Nomogram

Since LR's refer to actual test results before disease status is known, they are immediately more useful to clinicians than Sn or Sp.[55] Although the predictive use of LR's has limitations, LR's represent a distinct advantage over the traditional use of PPV and NPV. Likelihood ratios (LR+/LR-) and their respective 95% confidence intervals will be calculated for each diagnostic test and diagnostic test cluster assessed in this study.

2.3 Reference Criteria or "Gold Standards"

The traditional medical model of disease proposes that all disease may be defined by deviations from pathophysiologic norms and that the underlying cause of disease must be identified before appropriate corrective measures, in the form of treatment, can implemented.[59] Indeed, Taylor has stated that the current understanding of this model has come to be strictly associated with the identification of structural abnormality rather than referring to clinical or etiological events.[60] Although the simplicity of this model is attractive, it is well known that symptoms and pathology are not always strongly correlated in a number of conditions.[61-63]

Cervical radiculopathy is a condition in which the nerve root is insulted and typically results in pain, disturbance of function, and may often be accompanied by a variety of anatomic and pathophysiologic changes.[64pp.537-539] Therefore, CR is subsumed very well by the traditional medical model. It does not appear, however, that carpal tunnel syndrome is as well accounted for by this model as is CR. The term "syndrome" is defined as "a concurrence of symptoms" or "the aggregate of signs and symptoms associated with any morbid process and constituting together the picture of the disease".[65] Accordingly, a cluster of signs and symptoms may not necessarily be attributable to a distinct anatomical abnormality. Despite the connotations of the term "syndrome", the signs and symptoms of CTS are attributable to compression of the median nerve in the carpal canal.[66] Therefore, CTS may also be identified by a pathophsyiologic abnormality of the median nerve in a majority of patients.[67] Since both of the conditions of interest in this study may be defined on the basis of pathophysiological abnormalities, the ideal reference criterion (referred to hereafter as "gold standard") used to assess the efficacy of

diagnostic test procedures is one capable of detecting pathophysiologic abnormalities which may be singularly applied to a patient suspected of having either condition. An EMG/NCS examination that includes needle electromyography (EMG) and nerve conduction studies (NCS) of selected muscles and nerves of the patient's affected upper quarter meets this criteria and is commonly used in clinical practice to evaluate patients suspected of having either CR or CTS.[64pp558] A standardized EMG/NCS examination of the symptomatic upper quarter will be administered to all patients who participate in this study.. Abnormalities of the EMG/NCS examination will serve as the gold standard for the diagnosis of both conditions. The diagnostic properties of the EMG/NCS examination will now be discussed separately for each condition and a discussion of acceptable gold standards for patient outcome will follow.

2.3.1 Cervical Radiculopathy

The purpose of the EMG component of the EMG/NCS exam in patients with CR is to detect neural pathophysiology, specifically axonal-loss injury, and localize it to a cervical nerve root or roots.[64pp.548-555] The purpose of the NCS component of the EMG/NCS exam is to rule-out other causes of symptoms such as a diffuse peripheral neuropathy or more distal mononeuropathy [64pp541-545]

The typical EMG examination consists of assessing several limb muscles as well as the cervical paraspinal musculature. The selected limb muscles sampled each represent the integrity of the ventral primary rami of the 1 or 2 nerve roots from which they receive their innervation, which typically range from the C5 to T1 levels. The cervical paraspinal muscles are sampled at several vertebral levels with a needle electrode, which allows a general assessment of the dorsal primary ramus of the nerve root. EMG sampling of individual muscles consists of two main steps. First, the electrode is repeatedly inserted at several depths and in various directions in a given muscle in order to assess the integrity of innervation for a broad motor-unit territory. Normal muscle is electrically silent at rest. Therefore, during the examination the needle electrode is allowed to rest intermittently in the muscle in order to detect abnormal spontaneous electrical activity primarily in the form of fibrillations. Fibrillations and other less frequently observed forms of abnormal spontaneous activity occur in deinnervated muscle after 14-21 days

and are due to the development of acetylcholine hypersensitivity by the muscle membrane, which renders it unstable.[68,69] The second step is to have the patient voluntarily contract the muscle being examined which will elicit motor unit action potentials (MUAP). The morphology of the MUAP (amplitude, duration, and the number of phases) and recruitment pattern are then assessed for abnormalities.[64pp.548-555] Typical or standard NCS procedures used in the examination of patients with suspected CR are described below in the CTS section. The findings of the EMG and NCS examination are then integrated and if abnormalities are present and consistent with a lesion of the cervical root, then the diagnosis of CR is established.

Lacking a better method for detecting nerve root pathophysiology, investigators have attempted to establish the sensitivity and specificity of EMG by comparing it to other pathoanatomic procedures used in the evaluation and diagnosis of patients with cervical radiculopathy. These procedures include imaging studies (myelogram[70,71] and CT/myelogram[72]) and surgical observation.[70,73-75] Because of the difference in purpose of these procedures, the use of a pathoanatomic gold standard to determine the diagnostic accuracy of a procedure such as EMG, which defines abnormalities based on pathophysiology, is invalid.[6.64pp554] Instead, any comparison of the two procedures should merely be interpreted as a correlation or percent agreement and a degree of divergence would be expected. Even the use of surgical observation as a reference criterion for studies of EMG diagnostic accuracy is problematic and precludes the establishment of specificity because it could not feasibly or ethically be applied to the entire patient sample.[40] Determining the diagnostic accuracy of any test for spinal disorders is problematic due to the difficulty of establishing a suitable reference criterion or gold standard.[40] The diagnostic accuracy of needle EMG for cervical radiculopathy is no exception and depends upon the clinical parameter or reference criterion which is chosen.[40,76]

Needle electromyography is the oldest electrophysiologic examination procedure for the diagnosis of radiculopathy.[73] The percent agreement between positive EMG findings and surgically observed abnormalities in patients with CR ranges from 54% (95% confidence

interval (95CI)= 44%-64%)[75] to 100% (95%CI= 85%-100%)[73]; similar observations have been reported for patients with lumbo-sacral radiculopathies (78%-90%)[73,77,78] However, the interpretation of EMG findings and surgically observed abnormalities in patients with radiculopathy is somewhat confounding because some authors do not specify the criteria used to determine the presence of an abnormality: a herniated nucleus pulposus (HNP) or an irritated nerve root.[73,75,78] A diagnosis of CR based on the mere observation of an HNP is inappropriate.[45] Boden et. al. found herniations of the cervical intervertebral disc to be present in 18% of 63 asymptomatic volunteers[79] and an even larger percentage of false positive findings in the lumbar spine have been reported by numerous authors.[80-83] Wilbourn's statement that needle EMG is nearly 100% specific for the examination of patients with suspected radiculopathy[84] cannot be substantiated due to the methodologic limitations mentioned previously, but no one has reported a false positive EMG in patients treated surgically for CR;[70,73-75] the same cannot be said for imaging studies.[75,85]

Myelography is the imaging procedure EMG has been most frequently compared with. The correlation between myelography and EMG in patients with radiculopathy is consistently high for both the cervical (75% (95%CI= 61%-95%)[70,71] and lumbosacral (90% (95%CI= 77%-100%))[86] regions and a complementary relationship between the two procedures for the diagnosis of radiculopathy has been acknowledged in all identified reports.[77,86-91] The advantages of EMG versus myelography include: the ability to detect lateral root entrapment;[76] detect insidious disease processes;[75] and not injecting foreign material into the body.[64pp554] Shared disadvantages are that both procedures are invasive and involve various degrees of discomfort. A high percent agreement has also been reported when EMG is compared to CT in patients with lumbosacral radiculopathies (85%[77] to 89% (95%CI= 80-98%)[92]). In the only study in which data were statistically analyzed, EMG was found to be superior over CT ($P<.0001$) and the clinical exam for detecting which lumbosacral nerve root was involved.[77] The only study to compare EMG with CT in patients with suspected CR found an agreement of 67% for the two procedures (95%CI= 41%-93%)[93] Although the use of non-invasive imaging techniques such as CT and MRI for diagnosis of radiculopathy is appealing, both procedures support pathoanatomic diagnoses in 10% to 30% of the asymptomatic population depending on

age.[79-83,94,95] This is of concern because surgical intervention for patients with CR may often based on positive test results. Given the potential complications associated with surgery, the low morbidity associated with untreated CR, and the fact that prognosis for recovery is good in the majority of cases,[96-100] it can be argued that a diagnostic procedure which yields few false positive findings (i.e. is highly specific) is warranted for the diagnosis of patients with suspected CR.[101]

Other electrophysiologic examination procedures for the diagnosis of radiculopathy have been advocated in an attempt to increase the sensitivity of EMG and include an analysis of motor unit action potentials (MUAP) and the evaluation of evoked potential latencies (flexor carpi radialis (FCR) H-reflex, median and ulnar F-waves, and dermatomal somatosensory evoked potentials (DSEP)).[76] Although these additional procedures may increase the yield of abnormalities detected during the electrophysiologic examination,[75,93,102] muscle membrane instability observed during needle electromyography is still considered the hallmark diagnostic sign and the single most sensitive pathophysiologic method for establishing the diagnosis of both lumbar and cervical radiculopathy.[73,75,77,78,103]

There is some evidence that EMG may be useful in predicting the outcome of patients with radiculopathy. Two studies have reported that patients with normal pre-operative EMG findings have poorer surgical outcomes as expressed by symptom relief[75] or measured pain intensity compared to patients in whom pre-operative EMG abnormalities were observed ($p<.01$).[104] One study reported that patients with an abnormal pre-operative FCR H-reflex had a better clinical outcome at two years ($p<.03$) than did patients who had a normal pre-operative H-flex; a similar relationship was not observed for needle EMG findings.[70] Despite these reports, the relationship between EMG and patient outcome is still inconclusive because some studies used non-standardized outcome instruments with unknown psychometric properties,[70,75] the number of subjects or cell sizes were limited,[104] and data were not analyzed statistically.[75]

2.3.2 Carpal Tunnel Syndrome

Because NCS and needle EMG provide a unique way of directly assessing the integrity of sensory and motor nerve fibers, they have become the mainstay for the laboratory evaluation of CTS.[64p1396] The primary purpose of the EMG/NCS exam in patients with suspected CTS is to detect and localize abnormalities to the distribution of the median nerve. Additional purposes may be to rule-out other causes of symptoms such as a diffuse peripheral neuropathy, a more proximal mononeuropathy, and in some cases rule-out a concomitant CR.[64pp869-875,6] Usually one or more median innervated muscles is examined with needle EMG as well as a radial or ulnar innervated muscle for comparison, except in cases of suspected concomitant CR when more comprehensive muscle sampling is performed.[64pp873] The procedure for the EMG examination was described in section 2.3.1. In a typical NCS examination of a patient with suspected CTS, both the sensory and motor components of the median nerve are assessed. Surface electrodes are placed on the wrist or fingers to record evoked potentials when nerve stimulation at the wrist or elbow occurs. Alternatively, recording electrodes may be placed over the nerve at the wrist to record evoked potentials when the digit or palm is stimulated. For comparison, the ulnar nerve is examined in a similar fashion although the radial nerve may also be used.[105] The latency and amplitude of the evoked potentials are the most commonly assessed NCS parameters. The findings of the EMG and NCS examination are integrated and abnormalities of latency, amplitude and muscle membrane stability, when present and isolated to the median nerve distal to the wrist, help establish the diagnosis of CTS.[66]

In 1956, Simpson was the first to report the usefulness of median motor nerve conduction studies in the diagnosis of CTS.[106] His observations were later validated by a number of other investigators[107,108] and assessment of the sensory component of the median nerve was also included as advances in technology made this feasible.[109] Using intraoperative NCS, Brown confirmed that nerve conduction abnormalities of the median nerve in patients with CTS were localized to the area under the transverse carpal ligament.[110] Fullerton suggested that two mechanisms were responsible for the signs and symptoms of CTS: one is a rapidly reversible change in the nerve fiber associated with episodes of

ischemia and the other is due to slowly developing structural changes in the nerve fibers due to compression of the median nerve under the transverse carpal ligament.[111]

Clinical investigators have developed and refined a variety of techniques in order to maximize the sensitivity and specificity of NCS/EMG procedures for the diagnosis of CTS. These techniques include but are not limited to:[67,112] comparisons of latencies (bilateral median nerves and median nerve with ipsilateral ulnar and radial nerves); short segment mixed nerve latencies; sequential short segment (1cm) latencies; and comparison of nerve conduction velocity (NCV) across the carpal tunnel with NCV of a finger or forearm segment.[47,113,114] The reported specificity of sensory NCS is excellent. An assessment of long, short, and comparative sensory techniques as well as distal motor latency NCS in several large series of patients (n=100-300) suspected to have CTS is > .95 (95%CI= .95-1.0).[115-117] The sensitivity of NCS is lower and varies depending on the technique used: standard sensory conduction techniques range from .49[116] to .84[117]while short segment, mixed nerve conduction techniques; and techniques that compare the ipsilateral median and ulnar nerve range from .69[115] to .84.[117]

Two recent reports[47,113] show that the ratio of the NCV's across the carpal tunnel and NCV of either the forearm or digit is both sensitive and specific for the diagnosis of CTS. These later two works support the earlier findings by Kimura and Ayyar who tested 814 limbs and found slowing of the sensory NCV across the carpal tunnel relative to the forearm in 100% of CTS patients but not in any of the asymptomatic control subjects.[114] In a sample of 50 hands with clinically confirmed mild to moderate CTS and 40 normal controls, Padua et. al. computed the ratio of the NCV in the 3rd digit and the nerve conduction velocity across the carpal tunnel. This ratio was called the distoproximal ratio and reported a sensitivity of .98 (95%CI= 94-1.0) and a specificity of > .95 (95%CI= 95-100%) for the procedure.[113] Gunnarsson et. al. evaluated the diagnostic accuracy of a similar NCS technique in which the NCV across the carpal tunnel and the proximal NCV in the forearm is used to compute a NCV ratio. This ratio was obtained in 169 hands referred for neurophysiologic evaluation of CTS. The diagnosis of CTS was then retrospectively established three months later by using a combination gold standard of a

hand diagram, Symptom Severity and Functional Status Scale ratings, and standard NCS. If surgery was performed, relief of symptoms and observed median nerve abnormalities were required to establish the diagnosis of CTS. An receiver operating characteristic curve was used to determine the optimal sensitivity and specificity values which were .80 (95%CI= .68-.88) and.87 (95%CI= .76-.94), respectively.[47] The method reported by Padua et. al. will be used in this study due to its utility and ease of performance.[113]

Although NCS/EMG procedures are the only method of determining the physiologic status of neural elements (axon and myelin) in peripheral nerve, some patients have obtained relief after CTS surgery despite having a normal EMG/NCS examination.[118] Grundberg performed carpal tunnel release surgery in a series of 292 patients, thirty-two of whom were operated on despite normal nerve conduction studies and revealed the following: thirty one of the thirty-two patients experienced subsequent relief; no median nerve abnormalities were found in 22 of these patients; mild compression was observed in 8; and moderate compression was noted in one subject.[118] In addition, NCS abnormalities have been observed in asymptomatic subjects when certain NCS techniques are performed.[119] Most studies reporting on the false negative rate of EMG/NCS in the diagnosis of CTS have either been retrospective, have not provided a valid or unbiased reference criterion (i.e. "good" vs "poor" surgical or treatment outcome), or used less sensitive, long-segment NCS techniques.[67] Clinical opinion and the clinical examination have also served as gold standards for determining the diagnostic properties of NCS but the validity and reliability of these variables have not been well established. Magnetic-resonance imaging (MRI) of the carpal canal has also been considered for the diagnosis of CTS but its diagnostic and predictive value has yet to be determined.[120]

Similar to patients with CR, there is some evidence that NCS procedures may be useful for predicting outcome in patients with CTS who have been treated surgically; no reports were identified that dealt with non-surgically treated patients. Also similar is the quality and quantity of research related to the value of NCS procedures for patient prognosis in patients with surgically treated CTS: it is limited; the majority of studies are

retrospective;[121-125] few use standardized outcome instruments with valid psychometric properties;[121-127] and appropriate statistical analyses are usually lacking.[122.123.126.127] The one report that conducted a statistical analyses of the results was a retrospective study that assessed the outcome of 131 patients who underwent a second operation for CTS due to persistent or recurrent disabling symptoms. Pre-operatively, patients underwent a standardized EMG/NCS examination and completed the Brigham and Women's Hospital Hand Symptom Severity Scale (SSS) & Function Status Scale (FSS). NCS results were normal in 24 patients (18.3%) and abnormal in 107 patients (81.7%). Patients with preoperative NCS abnormalities had significantly better final SSS scores ($p<.005$) than patients with normal findings. The authors also reported that FSS scores were significantly improved at a p value of .07.[128] Several studies of poorer methodological quality have reported a relationship between abnormal preoperative NCS and improved post-surgical outcomes as measured by a variety of non-validated patients self-report instruments,[123.124] impairment measures,[122.123.125-127.129] and clinician or patient opinion[121.122.124] but conflicting reports also exist.[124.125.127.129]

In summary, neural impairment characterizes both CR and CTS and is a chief concern when managing patients with either of these two conditions.[130,6] Although NCS/EMG procedures are not 100% sensitive or specific for the diagnosis of CR or CTS, they are commonly used in the evaluation of patients suspected to have either condition. While there is debate regarding the precise role of NCS for the diagnosis of CTS,[66] the best objective diagnostic test continues to be an EMG/NCS examination.[64pp867] EMG/NCS procedures are the only way of assessing physiologic neural impairment in both conditions.[64pp. 554] In this study, the diagnosis of CR and CTS will be based on the presence of abnormalities of the EMG/NCS examination and opinion of the EMG/NCS provider. Once a diagnosis of either condition is established, the provider will classify the patient according to severity of EMG/NCS abnormalities as outlined in section 4.2.3, pp 75 -77. The CR and CTS classification schemes used in this study are modifications of similar, previously published classification systems.[127]

2.3.3 Patient Outcome

When deciding what variable or variables of interest will serve as a gold standard for patient outcome, the level of disablement, level of outcome, and the type of instruments or measures to be used are all issues that must be carefully considered.

Nagi's disablement model represents a traditional pathology-based approach to disability that is linear in nature.[131] However, numerous studies representative of a wide spectrum of medical care provide clear evidence that the relationship between the various levels of disablement depicted in Nagi's model are not always direct or proportional.[61,62,132-136] The presence of disease and impairment does not always result in functional limitations or disability and proportional changes in the severity of functional limitations or disability are not always observed.[61-63,70,104,122,123,132-134,137,138] This same relationship between the levels of disablement is also apparent in patients with functional limitation and/or disability; disease and impairment may be absent or not directly proportional to the severity of functional limitation and disability.[135,136] This non-linear relationship between levels of disablement has implications for deciding which outcome variables should be assessed when monitoring patient response to treatment in both clinical care and research settings.

Treatment intervention may be directed at different levels of disablement and may be assessed at a variety of outcome levels. A conceptual framework depicting the relationship between level of intervention and level of outcome has been proposed by Whyte which was adapted to fit Nagi's disablement model.[139] (Figure 5)

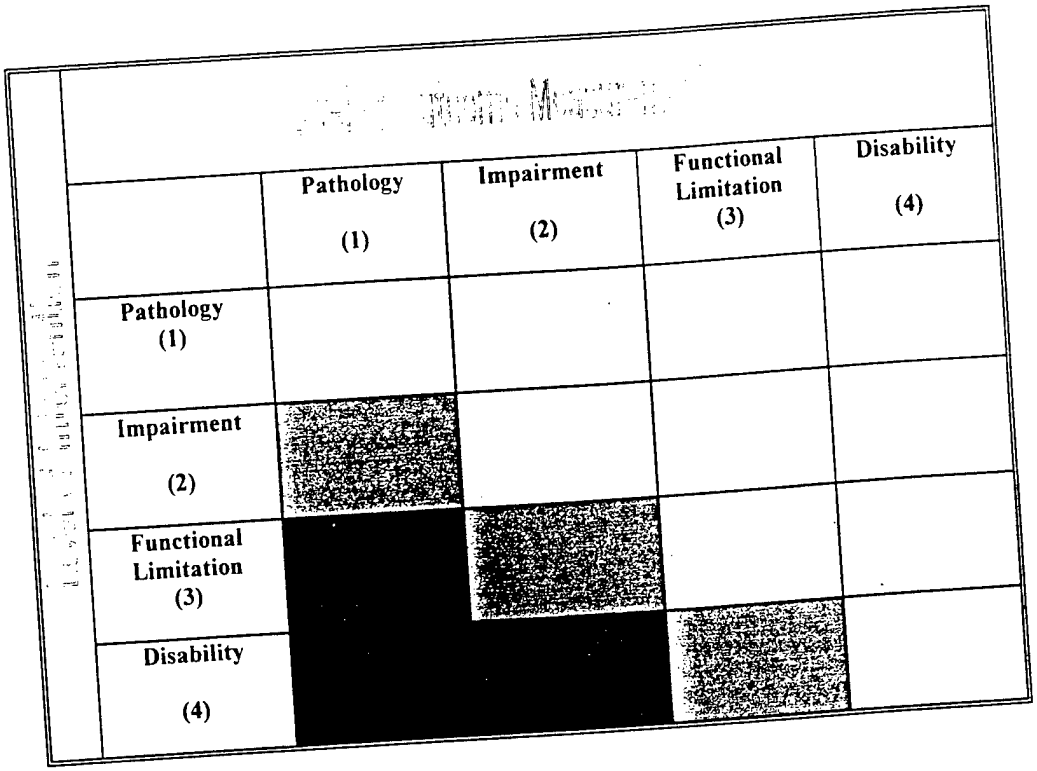


Figure 5 Whyte's Outcome Model (adapted from Whyte[139])

Because cells along the main diagonal represent congruent levels of intervention and outcome, it is thought that sensitivity to treatment effects is maximized because the treatment effects have a direct impact on the outcome of interest.[139,140] When levels of intervention and the corresponding levels of measurement of outcome are above or below the main diagonal cells, there is often a trade-off between measurement sensitivity (ability to detect change) and clinical relevance due to the impact of intervening variables.[139] Because the level of treatment and level of measured outcome may not always coincide, Whyte's framework may be helpful for choosing an outcome gold standard when conducting rehabilitation research. The interaction between the levels of intervention and levels of outcome must be considered in conjunction with the purpose of the outcome measurement.

Once level of intervention and level of outcome have been determined, it is still necessary to choose the particular instruments or measures that will be used to assess outcome. Many clinicians prefer to monitor a patient's change in status or outcome with

the same laboratory, imaging, physiologic tests, and clinical measurements that were used to diagnose the target disorder.[140] However, most of these variables are unsuitable for use as outcome measures because they primarily reflect pathology and impairment levels of disablement and are unresponsive to change even when an improvement in patient status has occurred.[63,70,132-134,137,141] In patients with radiculopathy, EMG abnormalities may be observed years after the initial onset of symptoms despite relatively minimal symptoms and disability.[70,104]. The same relationship is observed with NCS and patients with CTS: although marked improvement is observed in some patients following non-surgical and surgical intervention, NCS may show no improvement or remain abnormal.[122,123,127,137,138] Therefore, a change in the patient's level of functional limitation or disability often will not be detected if EMG/NCS findings (i.e. pathology/impairment level measurement) are used as an outcome measure in patients with CR and CTS. Although treatment will not be controlled, the outcome of interest in this study is primarily at the functional limitation and disability level.

Patient self-report instruments, health status assessment measures (HSAM) inparticular, have been shown to be valuable measures of patient outcome despite continued conceptual, methodological, practical, and attitudinal barriers to their use.[140] A number of HSAM whose purpose is to measure clinically meaningful change have established psychometric properties and often reflect the most relevant outcomes for patients and society.[140,142] Unfortunately, most HSAM are used for research purposes to compare groups of patients and what constitutes a meaningful clinically significant change score for a given instrument is largely unknown.[92,143-146] Other familiar indicators of outcome may be economic or related to the risk and complications associated with treatment of a given condition. These markers are familiar to many interested parties involved in the management of health care delivery but they frequently provide little or no useful information about the status of individual patients and are of little value for guiding treatment decisions.[140]

One problem for assessing the improvement or deterioration for qualities such as pain and function is that no objective external gold standard for such change exists.[143] In

addition, there is no consensus for what construct best serves as the gold standard for change.[92,145,146] Despite this limitation, several investigators have used either the patient's or clinician's ratings as a gold standard for outcome or change when assessing the responsiveness of health status assessment instruments.[92,143,146] Global perceived effect is an outcome measure for improvement that includes pain, functional status, and other constructs or dimensions that patients classify as important.[146] Jaeschke has described a 15pt global self-rating scale (GSRS)whereby patients may rate their own perception of improvement[143] The scale ranges from –7 ("a very great deal worse") to zero ("about the same") to +7 (a very great deal better). Intermittent descriptors of worsening are assigned values from –1 to –6 , and intermittent descriptors of improvement are assigned values from +1 to +6.

The complete list of descriptors with the corresponding values is as follows:

A very great deal worse	(-7)	About the same (0)		A very great deal better	(+7)
A great deal worse	(-6)			A great deal better	(+6)
Quite a bit worse	(-5)			Quite a bit better	(+5)
Moderately worse	(-4)			Moderately better	(+4)
Somewhat worse	(-3)			Somewhat better	(+3)
A little bit worse	(-2)			A little bit better	(+2)
A tiny bit worse	(-1)			A tiny bit better	(+1)
(almost the same)				(almost the same)	

The use of retrospective GSRS as a gold standard for change has been criticized because the reliability and validity of this method is unknown and patient recall of former health status may be inaccurate or biased.[147] Despite these potential limitations, the use of a retrospective global rating of change as an outcome gold standard represents a credible option in the absence of an external gold standard and continues to be a common, feasible, and useful method for assessing outcome.[143]

This study will measure two outcome variables. The first is the type of intervention, surgical or non-surgical, a patient received for his or her particular condition at 6 weeks from the time of enrollment in the study. This dichotomous grouping (surgical/non-surgical) takes into account economic considerations as well as concerns about morbidity and timeliness of intervention. If a patient is able to manage the symptoms of his or her

condition with non-surgical treatment, it may be possible to avoid the costs and complications associated with surgery.[6,15,98,99,121,148] Likewise, if a patient requires surgery in order to obtain symptom relief, less cost and debilitation may be experienced by offering the patient a surgical option sooner rather than later.[15,99,148,149] The second is patient perception of improvement using a GSRS of improvement. This global rating of improvement is the optimum outcome variable of choice in this study for several reasons: 1. It captures meaningful information representing several constructs that are of primary concern for the patient;[146] 2. Measures of neural and clinical impairment may be relatively unresponsive to change;[70,137] and 3. The MCID of region and disease specific HSAM for CR and CTS is unknown.

2.4 Critical Appraisal of Clinical Examination Measures and Self-Report Instruments for Cervical Radiculopathy and Carpal Tunnel Syndrome

An item or an item cluster from the clinical examination or patient self-report measures may be useful for diagnosis of condition and predicting patient outcome. However, most studies that have assessed the diagnostic test properties of items of the clinical examination and patient self-report instruments contain violations of many diagnostic test methodological research principles. Therefore the true diagnostic value of clinical examination items and self report instruments for CR and CTS is unclear or unknown. A critical appraisal of existing work related to items of the clinical examination and patients self-report instruments is discussed below.

2.4.1 Common Clinical Examination Measures

Several common clinical examination measures are used when evaluating patients with CR and patients with CTS.[16,34pp.50-52; 194-196.,150 pp.. ,151pp.120-126] However, the reliability and validity of many of these procedures has either not been reported or adequately established. Inter-rater reliability is a pre-requisite for establishing the diagnostic accuracy and predictive validity of any examination procedure used to measure impairment or aid in diagnosis.[24 pp.33-34] The inter-rater reliability, diagnostic accuracy, and predictive validity of the following commonly used clinical examination measures will be evaluated in this study:

Questions: In the majority of disease states, including neurologic disorders, an accurate diagnosis can be made from information obtained from the history alone over 75% of the time.[28-30] Although not common, a question regarding symptoms, symptom location and behavior, and history can be used and measured as diagnostic tests.[27] To be valid, diagnostic test questions must be developed using the same rigorous research methodology and metrics applied to other diagnostic procedures thought to be more objective, typically laboratory tests or clinical examination measures.[27] Diagnostic test questions are often the most powerful diagnostic measure for some conditions [28-30] and can possess sensitivity and specificity values of 1.0 in some cases.[49]

Certain questions and patient responses are thought to be of diagnostic value when examining patients with suspected CR[17,152] or suspected CTS[16] but have not been formally or only limitedly assessed. This study will assess the diagnostic properties of the questions and responses listed below which address symptoms frequently reported by patients with CR and CTS.[16,17,152] No data addressing the inter-rater reliability of these items are available. All 11 questions are listed on an evaluation form, which is contained in appendix C.

1. Which of the following symptoms are most bothersome for you?

   Pain

   Numbness & Tingling

   Loss of feeling

2. Where are your symptoms most bother some?

   Neck

   Shoulder or shoulder blade

   Arm above elbow

   Arm below elbow

   Hands and/or fingers

3. Which of the following best describes the **behavior** of your symptoms?

Constant

Intermittent

Variable (comes and goes)

4. Does your affected hand feel "fat" or "swollen"?

5. Do you have trouble with fumbling or dropping objects from your affected hand?

6. Does your entire affected limb and/or hand feel numb?

7. Do your symptoms **keep** you from falling asleep?

8. Do your symptoms **wake** you during the night?

9. Do your symptoms improve with moving or positioning your **neck**?

10. Do your symptoms improve with moving, "shaking", or positioning your **wrist or hands**?

11. Are your symptoms brought on or made worse when performing tasks that require a lot of grasping or finger movement?

Conventional Neurological Examination of the Upper Extremity: This examination includes testing of strength, muscle stretch reflexes (MSRs), and sensation testing. A standard neurological examination of the upper extremity is indicated for patients who present with radiating neck pain and symptoms of carpal tunnel syndrome.[16,153] Viikari-Juntura found moderate inter-rater reliability for sensory and strength testing (Kappa .40 - .64)[154] using standardized and operationally defined test procedures. Although the conventional neurologic examination is a standard component in the evaluation of patients with suspected CR and CTS, its value for the diagnosis of CR and CTS has not

been well established.[21] The reliability and validity of conventional neurological examination procedures of the upper extremity are summarized in Table 1. The reliability and validity of conventional neurological examination procedures specific for patients with suspected CTS are summarized in Table 2.

Table 1

Reported Reliability and Validity Coefficients for the Conventional Upper Extremity Neurologic Examination in Patients with CR

TEST	OPERATIONAL DEFINITION	RELIABILITY	VALIDITY
Sensation[154]	Patient response to light touch and pain. Representative dermatomal areas identified and test with fingers (light touch) and injection needle (pain). Three-level scale used: normal, hyperesthesia, hypesthesia	(range) Light Touch K= .41 - .62 Pain- K= .29 - .68	See Below
MMT[154]	Primarily based on side-side differences of selected muscles. Determination of bilateral deficit made on basis of age and general condition. Three level scale used: normal, reduced, markedly reduced	(range) K= .40 - .64 *%Agreement= 67 – 97[155] **Pearson r= .63 - .98[156]	See Below
MSR's[21]	Elicited from the biceps, brachioradialis, and triceps using a reflex hammer	Has not been reported	Gold Standard: Myelography 1 neurologic sign positive- Sn= .83 Sp= .70 2 neurologic signs or more positive- Sn= .62 Sp= .78

* Based on 5 point ordinal scale. Trunk and LE muscles also sampled
** Based on 12 point ordinal scale; status of subjects permitted only 5 possible levels. Lower extremity muscles also sampled

Table 2

Reported Reliability and Validity Coefficients for the Conventional Motor and Sensory Examination in Patients with CTS

TEST	OPERATIONAL DEFINITION	RELIABILITY	VALIDITY
Sensation[21,157-162]	Patient response to light touch and/or pain. Representative areas of the median nerve-field identified and tested with fingers (light touch) and pinwheel needle (pain). Three-level scale used: normal, hyperesthesia, hypesthesia	K= .41 - .62	(range) Sn .20 - .92 Sp .41 - .90
MMT[21,158-161,163]	Primarily based on side-side differences of abductor pollicus brevis or opponens pollicus. Determination of bilateral deficit made on basis of age and general condition. Three level scale used: normal, reduced, markedly reduced	K= .33 - .81	(range) Sn= .11 - .53 Sp= .80 - .92* See Below

*Extreme value of 1.0 has been reported[164]

Range-of-Motion and Wrist Diameter Measurements: Cervical range-of-motion (ROM) is frequently assessed when examining patients with complaints of neck pain [165] and cervical ROM measurements may be used as an indicator of treatment effectiveness or as a measure of patient outcome.[166] Cervical ROM is often impaired and may result in functional limitations in patients with cervical radiculopathy[6] and impaired cervical flexion ROM has been described as being useful for the diagnosis of CR.[167] Many schools of manual therapy hypothesize that certain patterns of restricted cervical ROM are indicative or pathognomonic of a particular underlying cervical disease or syndrome but no data have been collected to support this hypothesis.[168pp.50,92-106.169pp.173-174]

The cervical ROM measures obtained in this study include: flexion, extension, bilateral side-bending, and bilateral rotation. Intra-class correlation coefficient values of .84 - .92 have been reported for measuring cervical range-of-motion with a variety of devices but the cumbersome nature of most of these devices often limits their clinical applicability.[165.170.171] Studies reporting the reliability of cervical range-of-motion measurements taken with a bubble goniometer are limited. Lowery reported a Pearson r of .54 for inter-rater reliability of a cervical impairment measurement take with a bubble goniometer.[166] Hole and colleagues reported the inter-rater reliability for bubble goniometer measurements of cervical flexion/extension, lateral bending, and rotation as ICC coefficients (model 2,1) of .84, .82, and .81, respectively.[170] Documentation of measurement precision in the form of a standard error of measure is lacking for all measurement methods. The bubble goniometer method will be used to take all cervical ROM measurements in this study.

The wrist-ratio index is a proportion derived from the ratio of anterior-posterior (numerator) and medial-lateral (denominator) wrist width measured at the distal wrist crease in centimeters. Patients with a larger wrist-ratio are said to have a more "square" shaped wrist which is presumed to result in diminished carpal canal space or residual carpal canal volume and thus be a predisposing factor for CTS.[172.173] A ratio of >.7 is said to be predictive of CTS.[174] Johnson originally described the wrist ratio index in 1983 and reported a strong positive correlation between subjects who had a ratio >.70 and prolonged median distal sensory latency.[174] This positive correlation has been observed

by a number of other authors,[172.173.175] including one large prospective study of 665 consecutive presenting for evaluation of CTS.[173] In a sample of 228 patients, Kuhlman and Hennessy reported sensitivity and specificity values of .69 and .73, respectively, for the wrist-index when a ratio of ≥.70 was used to define a positive test.[172] Using the same threshold criterion of ≥.70, Gordon et. al. were able to predict the development of CTS in automobile workers over a 3 year period with a sensitivity of .74 and a specificity of .76.[175] In the one study that reported no correlation between the wrist-ratio index and NCS, the index was computed from measurements taken at the proximal wrist crease of asymptomatic rail-road maintenance workers. The extremely conservative NCS values used to establish the diagnosis of CTS in this study were inadequate for identifying mild cases of CTS, which would be expected to exist in an asymptomatic population (distal motor and sensory latencies of >4.4ms).[176] Although the wrist-ratio index appears to be useful for evaluating patients with suspected CTS, the reliability and measurement precision of this clinical measure is not known.

Provocation tests: Provocation tests are procedures designed to increase or decrease a patient's symptoms and usually have a dichotomous outcome. A positive test is thought indicate that the target disorder has a mechanical component and may be more responsive to treatment.[168pp.94.169.177pp.75-88] The basis for most of the provocative tests in this study is the fact that mechanical deformation (compression or tension) or alleviation of mechanical deformation (distraction or relaxation) of the neural elements increases or decreases, respectively, symptoms or severity of symptoms in patients with CR and CTS.[178-182] Mechanical deformation results in a reproduction or increase of the patient's symptoms due to ischemia and irritation of nerve axons whose firing threshold is elevated due to injury.[179.181]

Provocative tests for patients with CR may induce or alleviate mechanical deformation by the following mechanisms: enlargement or narrowing of the neural foramen;[180.183] peripheral neural elements placed on slack or stretch;[184.185] and an increase in intrathecal pressure.[150pp.123-127] Most provocative tests for patients with CTS employ external pressure directly[186.187] or indirectly[157] to the carpal tunnel that further increases the

already elevated pressure within the carpal canal [88.178.179] or utilize direct percussion of the nerve to excite hyperirritable or regenerating axons.[188.189]

Each provocative test assessed in this study is described below along with the operational definition of each test as it will be administered in this study.

1. Neck Compression Tests: Originally described by Spurling and Scoville as a test for cervical radiculopathy. In their description, the patient's head is laterally flexed towards the side of pain, and a compression force of ~15 lbs. is applied to the top of the head. A positive test is defined as the reproduction of the characteristic radicular pain.[183] Other authors have modified this test and incorporate neck extension and rotation towards the side of pain prior to applying a compression force of ~15 lbs to the head.[154,190]

Both test procedures will be used in this study and will be applied with the patient in a sitting position. The test as originally described by Spurling and Scoville will be designated Method A[183] and the modified version Method B.[190] **Method A** will be performed first and graded as positive or negative. Following the application and grading of Method A, **Method B** will then be performed and graded in the same manner.

2. Shoulder Abduction Test: The shoulder abduction test is performed on patients with complaints of radiating neck pain or radicular signs and symptoms.[184,191] Although the mechanism of pain relief is unclear, it is thought to be due to diminished tension on the irritable nerve root[184] A positive test is defined as the elimination of or decrease in symptoms

The Shoulder abduction test is shown in Figure 6. While sitting, the patient is instructed to place the hand of the affected extremity on the head in order to support the extremity in the scapular plane.[184] The test will be graded positive or negative.

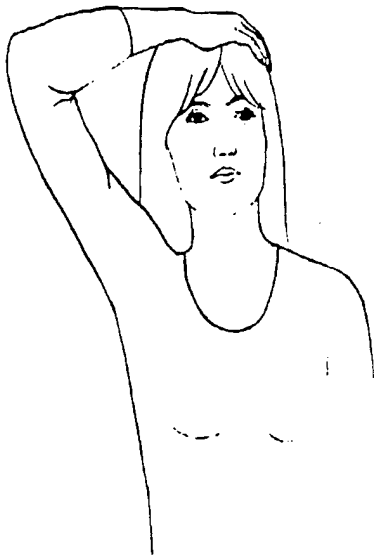


Figure 6

3. Valsalva Maneuver: The Valsalva maneuver is designed to detect a space-occupying lesion in the cervical spine, such as a herniated disk or an osteophyte. A positive test is defined as the reproduction or exacerbation of symptoms.[150pp.123-127]

While sitting, the patient is instructed to take a deep breath and hold the breath while attempting to exhale over a 2-3 second period with gradually increasing force. This test has been modified to include the gradual force build-up period. Because of associated morbidity, the Valsalva maneuver will not be performed by patients in this study who have cardiac disorders and patients with ophthalmic disorders other than visual acuity deficiencies.[192-194] The test will be graded positive or negative.

4. Distraction Test: The neck distraction test is performed on patients with complaints of radiating neck pain or radicular signs and symptoms.[150] A positive test is defined as the elimination of or decreased symptoms. If positive, a cervical disc herniation is suspected and indicates the potential for mechanical traction to be an effective treatment approach.

With the patient lying supine and the neck comfortably positioned, the rater will securely grasp the patient's head under the occiput and chin. An axial traction force, not to exceed ~30 lbs., will then be manually applied to the neck.[154] The test will be graded positive or negative.

5. Upper Limb Tension Tests (ULTT): The ULTT, or 'brachial plexus tension test' was originally described by Elvey in 1979.[195,185pp.577-585] Several modifications of Elvey's test designed to selectively stress the peripheral nerves of the upper extremity have since been proposed. Two basic ULTT procedures will be used in this study and are purported to emphasize tension in the median and radial nerve, respectively.[177pp.147-153]

ULTT A: With the patient supine and the cervical spine in neutral, the following motions will be sequentially applied to the symptomatic upper extremity and are illustrated in Fig 7 on the following page: 1) scapular depression (A), 2) gleno-humeral abduction (B), 3) forearm supination, wrist and finger extension (C), 4) shoulder external rotation (D), 5) elbow extension (E), and 6) contralateral then ipsilateral cervical lateral flexion (F). The patient is questioned regarding symptom reproduction throughout the maneuver. If symptoms are not reproduced during testing of the symptomatic limb, the test will then be applied to the opposite limb in an identical manner in order to compare elbow extension range-of-motion between limbs.[195,185pp.577-585] The test is considered positive in this study if the following conditions are met: 1) the test reproduces any portion of the patient's symptoms or pain complaints, 2) there are side-to-side differences in elbow extension when all previous motion sequences have been completed, and 3) for the symptomatic limb, contralateral neck lateral flexion increases symptoms or ipsilateral lateral neck flexion decreases symptoms. When a positive result occurs, the examiner will note and record the element in the range-of-motion test sequence (1 - 6) that elicited the positive result. The test is concluded when a positive result is obtained or when all motion sequences have been completed.

ULTT B: With the patient supine, shoulder abducted to 30^0, and the cervical spine in neutral, the following motions will be sequentially applied to the symptomatic upper

extremity and are illustrated in Fig 8: 1) scapular depression (A), 2) medial shoulder rotation (A), 3) full elbow extension ("locked") (B). 4) wrist and finger flexion (C), and 5) contralateral then ipsilateral cervical lateral flexion (as for ULTT A). The patient is questioned regarding symptom reproduction throughout maneuver. If symptoms are not reproduced during testing of the symptomatic limb, the test will then be applied to the opposite limb in an identical manner in order to compare wrist flexion range-of-motion between limbs. The test is considered positive if either of the following conditions are met: 1) the test reproduces any portion of the patient's symptoms or pain complaints, 2) there are side-to-side differences in wrist flexion when all previous motion sequences have been completed, and 3) for the symptomatic limb, contralateral neck lateral flexion increases symptoms or ipsilateral lateral neck flexion decreases symptoms. When a positive result occurs, the examiner will note and record the element in the range-of-motion test sequence (1 - 6) that elicited the positive result. The test is concluded when a positive result is obtained or when all motion sequences have been completed

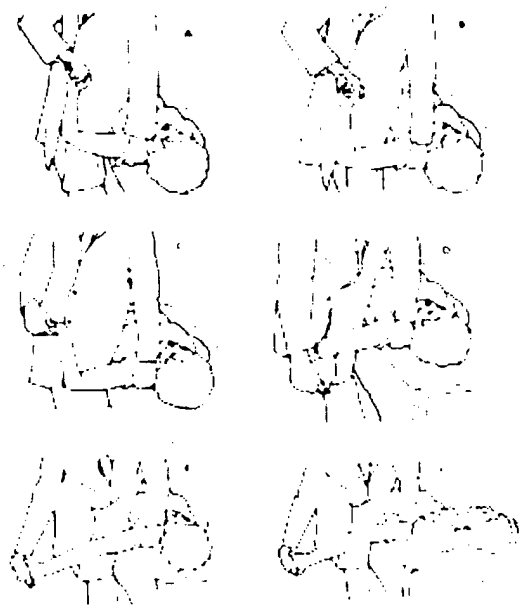


Figure 7 (Adapted from Butler[177])

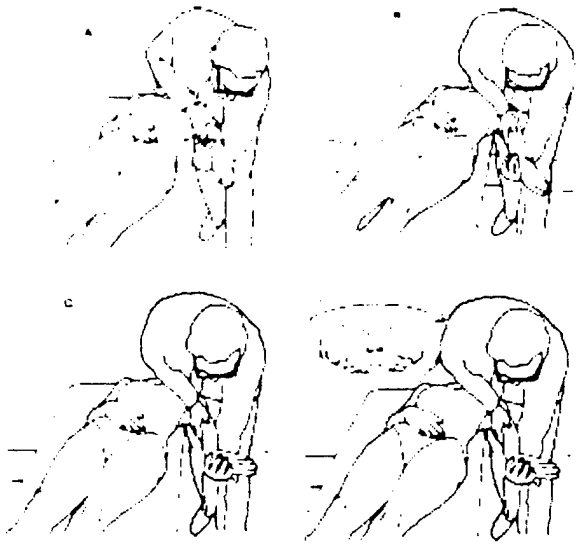


Figure 8 (Adapted from Butler[177])

6. Tinel's Sign: Two methods, classic (Method A) and provocative (Method B), will be performed. Both methods are reported to localize the level of a peripheral nerve injury by performing percussion directly over the nerve suspected to be involved.[196] The test was originally described as a method of detecting and monitoring nerve regeneration after laceration.[188] Others have described or applied the test as a provocative measure in order to reproduce the patient's symptoms.[16] In compression injuries such as carpal tunnel syndrome, regeneration will only occur after the syndrome has progressed to the point of nerve degeneration.[196] In the early stages of carpal tunnel syndrome, or in advanced cases in which the degeneration/regeneration process has reached a steady state, Tinel's Sign may be negative, even though the syndrome is present. In the case of suspected carpal tunnel syndrome, the median nerve is percussed over the carpal tunnel.[150]

Method A- With the patient sitting, elbow flexed 0-30^0, and the forearm in a supinated position, the patient's wrist and hand will be supported in a neutral position. A tendon reflex hammer positioned ~6 in. above the wrist will be allowed to fall 4-6 times over the median nerve located between the tendons of the flexor carpi radialis and the palmaris longus at the proximal wrist crease. A positive sign is considered to be present when the

patient reports a non-painful tingling sensation radiating distally along the course of the nerve.[188,196] The test will be graded positive or negative

Method B- The patient will be positioned as for method A above. Using a tendon reflex hammer, the Rater will directly percuss the median nerve located between the tendons of the flexor carpi radialis and the palmaris longus at the proximal wrist crease 4-6 times with mild to moderate force. A positive sign is considered to be present when the patient reports discomfort or pain at the wrist or radiating distally along the course of the nerve that is related to their condition. The test will be graded positive or negative.

7. Phalen's Test: Phalen's wrist flexion test was developed as a clinical test for carpal tunnel syndrome.[157] Maximal wrist flexion decreases the cross-sectional area of the carpal tunnel and compresses the median nerve between the flexor tendons and the transverse carpal ligament.[197] A positive test is defined as the reproduction or exacerbation of paresthesias or anesthesia in the cutaneous distribution of the median nerve in the hand.[157]

While sitting, the patient's elbow will be flexed $0\text{-}30^0$ and the forearm and wrist will be supported by the Rater in a pronated and neutral position, respectively. The patient's wrist will then be placed in a position of maximal flexion for a maximum of sixty-seconds.[157] For this study, the patient will be questioned with regard to symptoms at 15 second intervals during the sixty-second period.[198] The test will be graded positive or negative. The test is concluded when a positive test result is obtained or at the end of the maximum sixty-second time period.

8. Carpal Compression Test (CCT): The CCT was originally described by Durkan[186] as a clinical diagnostic test for carpal tunnel syndrome. The test is considered positive if the patient's symptoms in the cutaneous distribution of the median nerve are reproduced.

While sitting, the patient's elbow will be flexed $0\text{-}30^0$ and the forearm and wrist will be supported in a supinated and neutral position, respectively. Placing both both thumbs over the transverse carpal ligament, the rater will then apply a approximately 6 pounds of

pressure with each thumb. The pressure is maintained for a maximum of 30 seconds.[186] For this study, the patient will be questioned with regard to symptoms at 15 second intervals during the thirty-second period. The test will be graded positive or negative and is concluded when a positive test result is obtained or at the end of the maximum thirty-second time period.

The diagnostic properties and reliability coefficients of each provocative test, if known, are reported in Tables 3 and 4. Unlike tests of CR, a large number of studies of CTS have reported a wide range of diagnostic accuracy values for Phalen's test, Tinel's Sign, and the CCT (Table 4). One explanation for the variabity of these findings is errors in diagnostic test methodology. A summary of CTS provocative test studies and associated methodology is listed Figures 9 and 10

Table 3

Reported Reliability and Validity Coefficients for CR Provocative Tests

TEST PROCEDURE	RELIABILITY	VALIDITY
Spurling's[21.154]	(range) K= .61 - .71	Gold Standard: Myelography Sn= .36 Sp= .96
Shoulder Adbuction Test[21.154.184]	(range) K= .21 - .40	Gold Standard: Myelography Sn= .43 Sp= .80 % agreement= 68%
Valsalva Maneuver	Has not been reported	Has not been reported
Neck Distraction[21.154]	K= .50	Gold Standard: Myelography Sn= .40 Sp= 1.0
ULTT (may also used with CTS patients)[154]	K= .35	Has Not Been Reported

Table 4

Reported Reliability and Validity Coefficients for CTS ProvocationTests

TEST PROCEDURE	RELIABILITYS	VALIDITY#
Phalen's[16,157,160,162-164,186,199-213]		
	Intrarater K= .53	*Sn= .48 - .88
	Interrater K= .63	Sp= .32 - .90
Tinel's[16,157,160,162-164,186,199-201,203-214]	Intrarater K= .80	Sn= .25 - .74
	Interrater K= .79	Sp= .59 - .97
CCT[160,186,199-201,209,215,216 213,217]	Has not been reported	Sn= .21 – .89
		+Sp= .33 - .96

#Gold standard is NCS or NCS & compatible CTS symptoms in almost all cases (see Table 6 for detail)

$Reliability coefficients come from a single study[212]

*Extreme value of .11[164] has been reported

+Extreme value of .08[160] has been reported

TEST/*Study*	METHODOLOGICAL PRINCIPLE			
Phalen's and Tinel's	Clear Operational Definition?	Gold Standard	Spectrum?	Bias?
Borg et.al[163]	P=No; T=No	NCS	3; a	2*,3*
De Krom et. al.[160]	P=Yes; T-Yes	NCS/CTS Sx's	U; a	None
De Smet L et. al.[199]	P=Yes; T=No	NCS	U; a & b	2*,3*
Durkan JA[186]	P=Yes; T=N	NCS	3; b	2*,3*
Gellman H et. al.[204]	P=Yes; T=N	NCS/EMG	2; b	1,2*,3*
Golding DN et. al.[164]	P=No; T=No	NCS	3; c	2*,3*
Gonzalez J et. al.[200]	P=Yes;T=Yes	Clinical exam abnormalities and surgical relief	2; b	2*,3*
Heller L et.al.[205]	P=Yes; T=No	NCS/EMG	U; c	2*,3*
Katz JN et. al.[16]	P=Yes; T=Yes	NCS/EMG	U; c	None
Kuschner SH et. al.[206]				
Mossman SS et. al.[201]	P=No; T=Yes	NCS/EMG	3; unknown	1,2*,3*
++*Phalen GS*[157]	P=Yes; T=No	Clinical opinion	3;c	1,2,4
Novak CB et. al.[211]	P=Yes; T=Yes	Clinical Sx's	U; c	2*,3*
Rietz KA et. al.[208]				
Stewart JD et. al.(Tinel's only)[207]	No	NCS/Clincal Sx's	U; b	2*,3*
Seror P (Phalen's only)[202]	Yes	NCS	U; b	2*,3*
Seror P (Tinel's only)[218]	No	NCS/EMG	2	2*,3*
Szabo et. al.[213]	P=Yes; T= No	Surgical relief of symptoms	2; a,& b	2*,3*,4
Tetro MA[209]	P=Yes; T=No	NCS/EMG	U; b	2*,3*
Williams TM[210]	P=Yes; T=No	Clinical Sx's	U; b	2*
Gelmers HG (Tinel's only)[214]				

P= Phalen`s

T= Tinel's

*= Report did not exclude possibility

U= Unknown

++= Retrospective

Figure 9 (cont'd)

Types of Bias: 1= Work-up Bias

2= Diagnostic-Review

3= Test-Review

4= Incorporation

Spectrum: *Target Condition Severity* (stated or per NCS/EMG findings)-

1= Mild/moderate

2= Severe

3= 1 & 2 above

Control Group

a=Other or competing conditions, similar symptoms

b=Asymptomatic, "normal" subjects

c=None

Figure 9: Study Methodology for Phalen's Test (P) and Tinel's Sign (T)

TEST/*Study*	METHODOLOGICAL PRINCIPLE			
CCT	Clear Operational Definition?	Gold Standard	Spectrum?	Bias?
Durkan JA[186]	Yes	NCS	3; b	2*,3*
Durkan JA[215]	Yes	NCS/CTS Sx's	3; b	2*,3*
Gonzales J et. al.[200]	Yes	Clinical exam abnormalities and surgical relief	2; b	2*,3*
De Krom MC et. al.[160]	Yes	NCS/CTS Sx's	U; a	None
De Smit L et. al.[199]	No	NCS	U,a & b	2*,3*
Tetro MA et. al.[209]	Yes	NCS	U;b	2*,3*
Mossman[201]	No	NCS/EMG	3; unknown	1,2*,3*
Szabo et. al.[213]	Yes	Surgical relief of symptoms	2	2*,3*,4
Wainner RS et. al.[216]	Yes	NCS	3.c	2

*= Report did not exclude possibility

U=Unknown

Types of Bias: 1= Work-up Bias
2= Diagnostic-Review
3= Test-Review
4= Incorporation

Spectrum: *Target Condition Severity-*
1= Mild/moderate
2= Severe
3= 1 & 2 above

Control Group
a=Other or competing conditions, similar symptoms
b=Asymptomatic, "normal" subjects
c=None

Figure 10: Study Methodology for CCT

There is a dearth of research related to the efficacy of provocative tests for the diagnosis of CR. In contrast, a great deal of literature has been published regarding the efficacy of provocative tests for the diagnosis of CTS but due to the range of findings, few conclusions can be drawn. It is clear from the available evidence that the reliability and validity of commonly used provocative tests for the diagnosis of CR and CTS is not well established.

2.4.2 Common Patient Self-Report Measures

All but two of the patient self-report instruments assessed in this study are HSAM. Health status assessment measures can be used for three broad purposes which have been described as discriminative, predictive, and evaluative.[142] A *discriminative* instrument is used to distinguish between individuals or groups based on an underlying dimension without reference to a gold standard. A *predictive* instrument is used to classify individuals into distinct categories based on comparison with a gold standard in order to identify individuals who have or will develop a target condition or outcome. An *evaluative* instrument is used to assess clinically meaningful change over time[142]

An HSAM that demonstrated excellent psychometric properties for all three purposes mentioned above would be ideal. However, properties of an instrument that maximize one of the previously mentioned three purposes is likely to limit the ability of the instrument to fulfill the other two purposes well.[142] Several psychometric properties are essential before a HSAM or any patient self-report instrument can be meaningfully used for the patient management and include: reliability, validity, internal consistency, and responsiveness to clinical change.[219] The psychometric properties of the all the patient self-report instruments included in this study are acceptable and are discussed below.[220-223]

Clinicians have been reluctant to incorporate valid patient self-report instruments, in particular HSAM, into clinical practice despite the fact that they have been available for the last 20 years and are often more valid, reliable, and responsive than the traditionally used clinical examination measures of impairment.[144,224] A reason often given for this

reluctance is that they impose an excessive time burden on both the patient and practitioner.[144] Therefore, the use of patient self-report instruments in patient care should be done in a parsimonious fashion. If the self-report instruments used in this study are capable of fulfilling more than one purpose (discriminative, predictive, evaluative), then both clinician and respondent burden will be eased and it may facilitate more frequent use in clinical practice. Copies of all patient self-report instruments are located in appendix B.

The Neck Disability Index (NDI): The purpose of the NDI is to evaluate change over time in patients with neck pain. Vernon and Mior developed the NDI by modifying the Oswestry Low Back Pain Disability Index, which is a region-specific self-report measure of disability for patients with low back pain.[220] The authors first identified issues and activities considered most relevant to assessing the needs of patients with neck pain and submitted them to a group of clinical practitioners for review and consensus rating. The resulting items were then pilot tested in a group of 5 patients with whiplash injury. The final NDI consists of five items from the original Oswestry Index, two of which were revised considerably, and five new items. Seven of the items are related to activities of daily living, two are related to pain, and one item addresses concentration (ability to read). The original Oswestry Index format was retained and the terminology of the response statements were modified and made relevant for patients with neck pain. The six response statements are scaled from 0 to 5, with 0 representing no involvement or difficulty and 5 representing severe involvement or difficulty. The total NDI score is derived by summing the ratings of all 10 items so that a score of 0 represents good function while a score of 50 represents poor function. Riddle and Stratford successfully used an alternative scoring strategy that is similar to that of the Oswestry and accounts for items left blank by respondents. A percentage score is obtained by dividing the patients score by the maximum possible score for the number of items answered.[225] Although developed as an evaluative measure for patients with whiplash and chronic neck pain, the NDI has also been evaluated in patients with a wide variety of acute and chronic neck disorders.[225,226]

Vernon & Mior administered the NDI to 17 patients during an initial visit and then two days later, prior to the initiation of treatment. Test-retest reliability was reported as a Pearson r of .89 and internal consistency of the instrument was good (Chronbach's alpha=.80).[220] Binkley also found the NDI to have a high level of test-retest reliability (ICC=.89) when administering the instrument 3 days apart in a sample of 31 patients suffering from a variety of neck disorders.[226] Construct validity of the NDI is good and has been assessed in multiple settings using a variety of methods. Vernon and Mior found the NDI to be moderately correlated with a pain VAS (r=.60) and total scores from the McGill Pain questionnaire (r=.70).[220] In the study by Riddle et. al., the NDI was moderately correlated with clinician prognosis ratings (r=.66) as well as the physical (PCS; r=.53) and mental (MCS; r=.47) component summary scales of the Short-Form Health Survey (SF-36).[225] In addition, the NDI was more responsive than the PCS and MCS for detecting change in functional status between patients with different work status due to neck pain (altered vs not altered). Binkley reported a minimal level of detectable change for the NDI of 4.2 points.[226]

Brigham and Women's Hospital Hand Symptom Severity Scale (SSS) & Function Status Scale (FSS): The purpose of the SSS and FSS is to evaluate change over time in patients with CTS. The hand SSS and FSS were developed by Levine et. al. in 1993 as condition-specific scales to be used in the evaluation and assessment of outcome in patients with CTS.[221] The SSS consists of 11 statement items related to six domains said to be critical for the evaluation of CTS. These six domains, identified by a panel of hand surgeons, rheumatologists, and patients, include: pain; paresthesia; numbness; weakness; nocturnal symptoms; and over-all functional status. Each statement is rated by the patient on a 1 point (mildest) to five point (most severe) Likert scale. An overall SSS score is obtained by calculating the mean of the 11 individual items. A higher overall SSS score represents more severe symptoms and lower scores milder symptoms. The FSS consists of eight-items related to a variety of activities commonly performed by a broad spectrum of patients (i.e. young and elderly, workers inside and outside the home).

Each activity is rated by the patient on 1 (no difficulty) to 5 (cannot do at all) Likert scale. An overall FSS score is obtained by calculating the mean of the 8 individual items. A higher overall FSS score represents greater disability and and lower scores are representative of less disability.

The psychometric properties of both the SSS and FSS are acceptable. Levine et. al. assessed the test-retest reliability of both scales in a sample of 67 patients with confirmed CTS. Each scale was administered on two consecutive days and Pearson correlation coefficient's of .91 and .93 were computed for the SSS and FSS, respectively.[221] Because no universally accepted gold standard exists for measuring the severity of symptoms or functional status of the hand, scale validity was assessed by correlating the SSS and FSS scores with impairment measures.[142] It was hypothesized that more severe symptoms would be positively but weakly correlated with greater sensory and functional limitation measures. All correlations were in the expected direction and ranged from weak to modest. The following Spearman correlation coefficients were obtained: grip strength and pinch strength= .38 and .47 for the SSS and .50 and .60 for the FSS, respectively and two-point discrimination=.15 (SSS) and .42 (FSS). Internal consistency, as measured by Chronbach's alph, was reported to be .89 for the SSS and .91 for FSS. An effect size of 1.1 for the SSS and .71 for the FSS was obtained three months after surgery and indicated that both scales are sensitive to change in post-surgical CTS patients.

The psychometric properties reported by Levine et. al. for the SSS and FSS have been replicated by other authors in multiple clinical settings,[227] study designs,[227-229] and with patients receiving worker compenstation.[229] In addition, both the SSS and FSS have been shown to be more responsive than physical examination measures and generic or region specific patient self-report measures.[228]

Hand Diagram & 10-cm visual analogue scale (VAS): Pain drawing instruments or diagrams have proven useful for diagnostic and predictive purposes in patients with CTS,[16,223,230,231] patients with low back pain,[232,233] and are used primarily for psychological screening purposes. A pain diagram is usually administered by having the

patient mark the location of their symptoms on an anatomic diagram with symbolic markings descriptive of symptoms. The hand diagram developed by Katz and Stirrat for use in the evaluation of patients with CTS consists of anatomic images that depict the anterior and posterior surfaces of the left and right hands as well as the entire left and right upper extremities.[223] The hand images are located in the in the center of the page and bordered by the upper extremity images on each outside corner. In addition, a 10cm VAS for pain intensity is included at the bottom of the instrument. The 10cm pain VAS has been used extensively as an indicator of patient response to treatment and possesses construct validity,[234,235] is responsive to change,[236] and has excellent test-retest reliability (r=.99).[237] The original descriptors used by patients to complete the hand diagram were pain, numbness, tingling, and decreased sensation,[223] with a marking symbol peculiar to each descriptor. In a later study the authors collapsed the descriptors of numbness, tingling, and decreased sensation into a single response category because they frequently overlapped and it was difficult for some to distinguish the difference between these descriptors.[231] The diagram is graded by classifying the patient as having classic, probable, possible and unlikely CTS based the areas of the diagram that are marked.

In a sample of 63 patients treated for upper extremity paresthesias, the sensitivity and specificity of the Hand Diagram was reported to be .80 and .90, respectively.[223] Intra and Inter-rater reliability was determined for 54 randomly selected Hand Diagrams and reported as a percent agreement of 91% and 84%, respectively.[223] This initial report was retrospective and CTS prevalence was 88%, which limited conclusions about the validity of the Hand Diagram. However, other large, prospective trials assessing the diagnostic accuracy of the Hand Diagram alone and in combination with other diagnostic tests have reported sensitivities that range from .61[16] to .64 [230] and specificities that range from and .71[16] to .73[230], respectively. There is limited evidence that the hand diagram may be useful for prognosis in patients with CTS who are treated surgically.[231] Later studies included patients with a variety of upper extremity disorders, workers compensation cases, and are of stronger methodological quality. In summary, the Hand Diagram appears to be a useful self-report instrument for the diagnosis of CTS and may be useful for predicting outcome in surgically treated patients.

The Fear Avoidance Behavior Questionnaire (FABQ): Lethem and Slade et. al. proposed the Fear-Avoidance Model of Exaggerated Pain Perception which is based on data concerning the pain coping strategies used by patients.[238.239] The model proposes that an individual's response to pain can be described on a continuum: from a minimal fear of the painful symptoms and motivation to continue normal activities to the greatest extent possible, to a strong fear of painful symptoms and an avoidance of painful activities. Patients who respond in the former manner may be described as "confronters" and those in the latter manner as "avoiders". Confronters will tend to rehabilitate themselves while avoiders become increasingly deconditioned and disabled as a result of their avoidance behavior.

The FABQ is a self-report measure developed by Waddell et. al. in order to measure the fear-avoidance beliefs of low back pain patients.[222] The FABQ consists of 16 items and has a two factor structure. One factor concerns fear-avoidance beliefs related to work (11 items) and the other factor concerns fear-avoidance beliefs related to general physical activity (five items). The FABQ has been demonstrated to have acceptable psychometric properties.[222]

In a prospective study of 300 patients with acute low back pain, Klenerman et. al. found the FABQ and several other indicators of fear-avoidance beliefs to be the best predictors of which patients condition would become chronic.[240] Fritz et. al. found the work FABQ to be the best predictor of return to work at four weeks for a group of 67 patients with occupationally related acute LBP: sensitivity was perfect Sn (1.0) and Sp was.63 when a cut-off score of 30 is used (LR+= 2.7, LR-=.02).[241] The FABQ has not been used to predict chronicity for patients with CR or CTS. The use of the FABQ for patients with CR and CTS in this study is considered acceptable for the following reasons: the fear-avoidance model is based on the coping strategies of patients with a variety of conditions;[238.239] the FABQ has acceptable psychometric properties;[222] and the FABQ assesses factors that would not be limited to patients with LBP (i.e. fear of pain related to work and general physical activity)[238.239]

3.0 RESEARCH HYPOTHESES

1. Reliability: the clinical examination variables listed below will demonstrate the following specified levels of reliability[242.243p514]

a. **Excellent** (K > .75 or ICC >.90)

CTS Variables	CR Variables
Questions: 1.- "Most bothersome symptoms..",	1.- "Most bothersome symptoms..",
2.- "Where most bothersome.."	2.- "Where most bothersome.."
3.- "Symptom behavior.."	3.- "Symptom behavior.."
4.- "Hand fat/swollen.."	6.- "Entire limb numb.."
5.- "Fumbling/dropping.."	7.- "Symptoms keep from sleep.."
6.- "Entire limb numb.."	9.- "Neck movement improves.."
8.- "Night symptoms wake.."	
10.- "Hand shaking improves.."	
11.- "Worse with hand use.."	
Wrist Ratio	Neck ROM
Tinel's A & B	

b. **Fair to Good** (K= .40 -- .75) or **Good** (ICC= .75 -- .90)

CTS Variables	CR Variables	
Phalen's	Spurling's A & B	Sensation
CCT	Shoulder Abduction	MMT
Sensation	Valsalva	MSR's
MMT	Neck Distraction	

c. **Poor** (K < .40) or Poor to Moderate (ICC <.75)

CTS Variables	CR Variables
ULTT A & B	ULTT A & B

2. Test Diagnostic Accuracy: the concurrent validity of the clinical examination variables listed below will be determined to be acceptable or unacceptable based on the following criteria.[54,55]

a. **Acceptable** (Sn or Sp ≥ .70 or LR+ ≥ 2.0 or LR- ≤ .50)

CTS Variables	CR Variables	
Questions 2,4,5,8,10, and 11	Questions 2 and 7	Valsalva
Phalen's	Sensation	Neck Distraction
Tinel's A and B	MMT	(Spurling's A and B
CCT	MSR	Shoulder Abduction
Wrist Ratio		
Hand Diagram		

b. **Unacceptable** (criteria for acceptability not met)

CTS Variables		CR Variables
Questions 1,3, and 6	ULTT A and B	Questions 1,3, 6 and 9
Sensation	10cm VAS	ULTT A and B
MMT		Neck ROM
SSS & FSS		NDI
		10cm VAS

3. Test Predictive Validity: As with test diagnostic accuracy, the predictive validity of the clinical examination variables listed below will be determined to be acceptable or unacceptable based on the following criteria..[54,55]

a. **Acceptable** (Sn or Sp ≥ .70 or LR+ ≥ 2.0 or LR- ≤ .50)

CTS Variables	CR Variables
Question 6 and 10	Questions 6 and 9
Wrist Ratio	Neck Distraction
Hand Diagram	Shoulder Abduction

EMG/NCS	EMG/NCS
FABQ	FABQ

b. **Unacceptable** (criteria for acceptability not met)

CTS Variables	CR Variables
Questions 1-5,8,10,11	Questions 1-3, and 7
Sensation	Sensation
MMT	MMT
Tinel's A and B	Neck ROM
CCT	Spurling's A and B
Phalen's	Valsalva
ULTT A and B	ULTT A and B
10cm VAS	NDI
SSS & FSS	10cm VAS

4. Test Item Cluster (TIC):

a. It is hypothesized that for both CTS and CR, a combination of clinical examination variables and/or patient self-report items can be identified that yield acceptable levels of diagnostic accuracy (based on previous definition of acceptable).

b. It is hypothesized that for both CTS and CR, a combination of clinical examination variables, and/or patient self-report items, and/or EMG/NCS findings, can be identified that yield acceptable levels of predictive validity for type of intervention and patient perception of improvement. (based on previous definition of acceptable).

If acceptable test inter-rater reliability, diagnostic accuracy, and predictive validity is established for the clinical examination and patient self-report measures considered in this study, the benefits realized include but are not limited to: interpretable test results; more accurate clinical decision making with regard to diagnosis and treatment; more accurate estimation of patient prognosis; and a substantial reduction in medical costs and patient discomfort. Acceptable test reliability and validity will allow further research of

the predictive validity of these techniques, permit their wide application in patient outcomes research, and allow their confident application in clinical practice.

4.0 RESEARCH DESIGN AND METHODS

This was a multi-center, prospective, descriptive study designed to quantify the reliability, diagnostic accuracy, and predictive validity of commonly used clinical examination and patient self-report measures used to diagnose patients with suspected CR and CTS.

4.1 Inclusion Criteria

To be eligible for participation in this study, individuals must have been aged 18 - 60 years and referred for EMG/NCS testing to rule-out CR and/or CTS. Only patients judged by the EMG lab evaluating physician to have signs and symptoms compatible with cervical radiculopathy or CTS were eligible to participate. In addition, the patient's current episode of symptoms was required to exceed 4 weeks but not 12 and 24 months duration for CR and CTS, respectively. Patients with the following conditions were disqualified from study participation:

1. Systemic disease known to cause a generalized peripheral neuropathy.
2. Primary complaint of bilateral radiating arm pain
3. History of conditions involving the affected upper extremity which might adversely affect the individual's level of function
4. Off work for >6 months due to the condition.
5. Previous history of surgical procedures for pathologies giving rise to neck pain or for CTS
6. Patients who have had previous EMG/NCS testing of their symptomatic limb for CR and/or CTS.

All consecutive patients referred to the EMG laboratories of both Montefiore and Presbyterian Hospital for EMG/NCS testing to rule-out CR and/or CTS received information about the study and complete a study screening form (appendix A). If, after

examining the patient and reviewing the screening form, the EMG lab provider determined if the patient was eligible for study participation, the patient was asked by the provider to participate in the study. Prior to obtaining informed consent, the study investigators or their representative at distant participating sites explained the study in detail to the subject. If the patient agreed and gave informed consent in compliance with the standards of the Biomedical Internal Review Board of the University of Pittsburgh or the Internal Review Board of the other respective participating facilities, he or she was admitted into the study. Volunteers were also be recruited from the EMG laboratories of the following participating Military Treatment Facilities: The National Naval Medical Center (NNMC), Bethesda, MD; Wilford Hall USAF Medical Center (WHMC), San Antonio, TX; and the Air Force Academy Hospital (AFAH), Colorado Springs, CO.

4.2 Methods

4.2.1 Procedure

A video tape of all clinical examination procedures and the disto-proximal NCS technique as well as a clinical examination handbook and an EMG handbook detailing the performance of each clinical examination measure, equipment settings, and EMG/NCS procedures was distributed to each participating center prior to data collection. All physical therapist raters at each participating facility viewed the tape and read the clinical examination handbook in order to familiarize themselves with the clinical examination measures. In addition, all raters participated in at least two practice sessions during which all clinical examination measures, except the asking of questions, were performed. Physical therapist raters practiced applying the specified amount of compression or distraction force required for the Spurling test, distraction test, and CCT using a bathroom scale, mechanical traction device, and pinch gauge, respectively. All EMG providers viewed the tape and read the EMG/NCS handbook in order to familiarize themselves with the disto-proximal NCS procedure, EMG/NCS equipment parameters, and procedure protocol.

Once a patient was determined to be eligible and agreeded to participate in the study, the patient underwent a standardized EMG/NCS examination of the affected upper quarter

and completed the following self-report instruments which are listed in appendix B: NDI, FSS, SSS, FABQ, VAS, and Hand Diagram. The electrophysiologic examination consisted of established EMG/NCS procedures.[67,112,64pp.541-555] All EMG/NCS testing was performed by a physician, physical therapist or evoked potential technician with electrophysiologic testing credentials.

Within one week following the standardized EMG/NCS examination, the patient underwent two standardized clinical examinations administered by two physical therapist raters. The second examination was required in order to determine the reliability of the clinical examination measures used in this study. Therapist raters 1 & 2 were blinded to the patient's diagnosis or suspected condition. Rater 1 obtained responses to the 11 questions related to the patient's symptoms and performed the clinical examination measures with all participating patients. If any reproduction or increase in the patient's symptoms occurred, the therapist allowed the symptoms to return to baseline before administering the next test procedure. Each clinical examination procedure was graded or interpreted as previously described. Following a five-minute rest period, a second rater (rater 2) re-administerd the clinical examination measures to the patient in an identical manner. The 11 provocation tests were administered in alternating order with each new patient to control for order effects. Rater 2 did not administer the 11 questions of history to the patient prior to the examination but obtained patient responses 2-3 days following the examination. The 11 questions of history were administered to the patient by Rater 2 at the next follow-up visit or by telephone. The delay in obtaining responses to questions of history by Rater 2 was required to prevent item recall by the patient, which could confound the interpretation of reliability. The clinical examination results obtained by the first PT rater were used for all computations of diagnostic test accuracy.

All patients were mailed a 15-point GRCS six-weeks from the date of their initial clinical examination and were asked to rate their improvement. Patients were also mailed a treatment form and asked to document whether they had surgical intervention and list all non-operative treatment interventions they had received since their initial examination. All clinical evaluation forms are contained in appendix C.

4.2.2 Patient Demographic Data and Past Medical History

The following demographic data was collected: age, gender, specialty from which referred, workers compensation and litigation status, and employment status. Past medical history data was collected and included any previous or existing medical conditions, risk factors for generalized peripheral neuropathy, information related to the onset and duration of the current episode of symptoms, and whether or not the patient has had previous evaluation and treatment for the current condition. The EMG/NCS provider documented his or her suspected diagnosis for the patient as well as the diagnosis suspected by the referring provider. In addition, the EMG/NCS provider reviewed the patient's medical record and documented the findings of any available imaging studies, prescribed medication, and conservative treatments related to the patient's condition.

4.2.3 Standardized Electrophysiologic Examination

Surprisingly, little has been published to document the reliability of either standard NCS measurements (latency, velocity, amplitude) for the median and ulnar nerve[244] or the needle EMG examination. Two studies that used analysis of variance and paired t-tests found no differences in latency means in test-retest studies[244-246] but this approach is inadequate for establishing reliability.[247] In one recent unpublished study, Moore et al. found excellent intrarater reliability within a single measurement session for both distal sensory latencies (DSL) (ICC 2,1 = 0.98) and distal motor latencies (DML) (ICC 2,1 = 0.98).[248]

A reliability coefficient has not been reported for the intra or inter-rater reliability of the needle EMG examination. However, needle EMG is the most sensitive electrophysiologic procedure for detecting axonal loss occuring in cervical and lumbar radiculopathies[73.88] and multiple studies have documented its strong positive association with myelography, computed tomography, and surgical findings (percent agreement= 75%, (95CI 61% 95%)[70.71], 89%, (95CI 80%-98%)[249], and 78%-90%[73.73.77.77.78] respectively). Because of its strong association with multiple other diagnostic studies and surgical observation, the reliability of the EMG examination may be considered acceptable.

Following a history and neuromusculoskeletal screening examination, the EMG/NCS tester (Tester) thoroughly explained all EMG/NCS testing procedures to the patient and answered any questions. The patient was asked to lie supine on an examination table with the symptomatic limb toward the Tester. The temperature of the limb to be tested was assessed using a standard surface thermistor placed in the palm of the hand to be tested at the level of the metacarpal head. Hand temperature was $>32^0$ C prior to NCS testing. The area over which the electrodes were placed will was cleansed with an alcohol swab in order to decrease skin impedence. If the patient's hand temperature was $< 32^0$ C, the hand was placed in water warmed to 34 - 40^0 C and reassessed until hand temperature reached the acceptable limit.[64pp.29-64]

All EMG/NCS units had a current equipment safety rating prior to use. The instrument settings listed below were used as default parameters for the respective test procedures. Equipment settings were changed in order to obtain clear and interpretable test responses when technical difficulties were encountered. Any changes made to default parameters during testing were documented.

Electromyograph Instrument Parameters

A. Orthodromic evoked SNAP/CNAP:
1) Gain:
SNAP 20uV/division,
CNAP 20-50uV/division
2) Sweep speed: 1ms/division
3) Filter: 20 - 5000 Hz
4) Stimulus duration: .1 ms

B. Evoked CMAP
1) Gain: 2mV/division
2) Sweep speed: 2ms/division
3) Filter: 20 - 10,000 Hz
4) Stimulus duration: .1 ms

C. H-Reflex
1) Gain: 500 - 1,000uV/division
2) Sweep speed: 5ms/division
3) Filter: 20 – 10,000 Hz
4) Stimulus duration: .5 ms

D. EMG
1) Gain:
Insertional &
spontaneous- 50 -100uV/division
Recruitment- 1,000uV/division
2) Sweep speed:
Insertional &
spontaneous-10ms/division
Recruitment- 10ms/division and
100ms/division
3) Filter: 20 - 10,000 Hz

Commercially available tape, conductive gel, surface bar, and surface disc electrodes were used for nerve conduction studies. All electrode surfaces were wiped with alcohol between patients. Commercially available disposable 40mm or 50mm monopolar needle electrodes were used for all EMG testing. Used electrodes were disposed of in receptacles designated and approved for such use (i.e. sharps bucket).

Nerve conduction studies were performed first, followed by needle electromyography. All distances used for electrode placement and to calculate NCV were measured along the anatomic course of the nerve with a tape measure and recorded in millimeters.

Nerve Conduction Procedures: The stimulator was set to zero prior to each nerve conduction test. For each nerve study, the patient was notified prior to nerve stimulation. Several stimuli of gradually increasing intensity were delivered until a maximal response was obtained.[64pp.29-64] Evoked response parameters were measured and recorded for each response in the following manner:

1. Amplitude (microvolts) - peak-to-peak for SNAP/CNAP responses and baseline-to-peak for motor responses
2. Latency (milliseconds)- peak for SNAP/CNAP responses and departure from baseline for motor responses.
3. Nerve conduction velocity (M/s)- NCV is the quotient obtained by dividing the nerve segment distance in millimeters by the relevant nerve segment latency in milliseconds. For median and ulnar motor nerves, the relevant nerve segment latency is first obtained by subtracting the distal motor latency from the proximal motor latency.

The following NCS protocol was performed in order and in a standard, previously reported fashion.[113.250,64pp.29-64]

1. Median and ulnar nerve orthodromic palmar CNAP @ 8.0cm (latency & amplitude)
2. Median SNAP distal-proximal ratio (latencies & NCV's):

   Stimulation site: Ring electrodes placed on the third digit (D3), cathode proximal
   Recording site 1: **midpalm-** After obtaining an evoked potential the ring to D3 latency is recorded. Next, the NCV for the distal segment (D3 to palm) is calculated by dividing the measured distance by the latency.
   Recording site 2: **proximal wrist crease-** After obtaining an evoked potential, the proximal segment latency (midpalm to proximal wrist crease) is obtained by subtracting the distal segment latency (obtained in the previous step) from the ring to prox wrist crease latency. NCV for the midpalm to proximal wrist crease segment latency is then calculated. This is done by subtracting the distal latency (midpalm to D3) from the proximal latency (midpalm to proximal wrist crease). The NCV for the proximal segment is then calculated by dividing the measured

midpalm to proximal wrist crease segment distance by the calculated latency for that segment.

Calculate the distal-proximal ratio: Divide the NCV of the D3 to mid-palm segment by the NCV of the mid-palm to proximal wrist crease segment to obtain a proportion.

3. *Ulnar orthodromic SNAP @ 14cm (latency & amplitude)- done only if palmar CNAP is prolonged or technically unobtainable

4. Median and Ulnar distal CMAP @ 8cm (latency & amplitude)

5. Median and Ulnar NCV (latency, amplitude, and NCV)

Forearm segment- median and ulnar nerve

Elbow segment- ulnar nerve

6 *Median and Ulnar F-wave (latency)- done only in the absence of motor latency abnormalities or NCV abnormalities for each respective nerve.

7. H-Reflex- record flexor carpii radialis affected side (latency)

8. *If the median orthodromic CNAP or median distal-proximal NCV ratio is abnormal, then step 1) and 2) will be repeated on the opposite hand. If the median study of the asymptomatic side is abnormal, then step 4) will be repeated.

9. H-Reflex- record flexor carpi radialis opposite side (latency)

*Conditional procedures

Needle Electromyography Procedures: The skin of the limb to be sampled was cleansed with an alcohol wipe prior to needle electrode insertion. Each of the following muscles was examined for insertional, spontaneous, and recruitment activity in the following manner: mid and lower cervical paravertebral muscles, deltoid, triceps brachii, extensor carpi radialis longus/brevis, flexor carpi radialis, abductor pollicus brevis, and first dorsal interrosseus.

1. Insertional activity- Observed and recorded as normal, increased, decreased, for each muscle sampled.

2. Spontaneous activity- For each muscle site sampled, the tester utilized the standard quadrant/level method for a total of 12 observations at each sampling site.[64pp.29-64] Care was taken so that no electrode movement occured when making a determination of the

presence or absence of spontaneous activity. Spontaneous activity in the form of fibrillations and positive sharp waves (PSW) was graded 1+, 2+, 3+, or 4+. The presence and type of other spontaneous wave-form activity was documented as appropriate.

3. Motor unit analysis- Motor unit action-potential (MUAP) activity, consisting of recruitment and morphology, was assessed at least once for each limb muscle site sampled at a gain of both 100uv and 1,000uv; MUAP activity of paracervical muscles was not be assessed. The recruitment frequency/number (ratio) method was used to assess MUAP recruitment. The assessment of MUAP morphology was made when rise time was maximal (<500us) and included both number of phases and amplitude. Motor unit morphology was graded as **normal** or **increased** polyphasic, and motor unit amplitude was graded as **normal**, **increased**, or **decreased**.

Additional Procedures: Other EMG/NCS procedures or additional muscle sampling were performed as indicated from the clinical examination and were based on the Tester's opinion. The EMG provider documented all additional EMG/NCS procedures performed and/or additional muscles sampled.

Grading & Interpretation: For NCS procedures, the following previously established normal values listed in Table 2 were used in this study :

Table 5: Type of NCS Studies and Associated Normal Values

STUDY	PARAMETERS		
	Latency (ms)	Amplitude (uV)	NCV (M/s)
1. Median and Ulnar midpalmar CNAP @ 8.0cm.	≤2.2 or <3 med/uln diff	Med ≥40 Uln ≥11	-- --
2. Median SNAP distal-proximal ratio: a. 3rd digit to midpalm (NCV) b. midpalm to wrist (NCV) (10cm separation of midpalm and wrist cathode recording sites) c. Calculate distal-proximal NCV ratio by dividing a. NCV by b. NCV	--	--	Ratio <1.0
3. Ulnar SNAP @ 14cm.	<3.7	>12	--
4. Median and Ulnar distal CMAP	>4.3 Med >3.6 Uln	≥5000 ≥5000	--
5. Median and Ulnar F-wave	>32.0 or ≤.5ms med/uln diff	--	--
6. Median and Ulnar NCV median and ulnar nerve in forearm ulnar nerve across elbow	--	≥5000; <20% drop from prox/dist. stim site	≥50
7. H-Reflex- flexor carpii radialis	<19.0 or ≤1.0ms R/L diff or ≤calculated latency (.29 + .1905(arm length cm)±.84)	--	--

Each NCS was graded as abnormal if it exceeded normal values for that study.

For needle EMG procedures, the following previously established criteria were used to grade insertional activity, spontaneous activity (fibrillations and PSWs), and MUAPs:

1. Insertional Activity:

   Normal- electrical activity persists no longer than 50ms following cessation of needle electrode movement.

   Increased- electrical activity persists longer than 50ms following cessation of needle electrode movement

   Decreased- few if any electrical potential detected during or following needle electrode movement

2. Spontaneous Activity: Graded in accordance with Figure 11 below.

Figure 11: Spontaneous Activity Characteristics and Grading

Grading	Characteristics
0	No fibrillations or PSW
1+	Persistent/unsustained single trains in at least two sites of muscle sampled
2+	Moderate numbers in three or more sites of muscle sampled
3+	Many in all muscle sites sampled
4+	Baseline obliterated with fibrillation potentials in all muscle sites sampled

Other types of spontaneous activity consistent with denervation, when observed, were documented.

3. Motor Unit Action Potential (MUAP) Recruitment Analysis:

   Normal- frequency of preceding motor unit 5-10Hz prior to recruitment of a successive motor unit.

   Decreased- Ratio of fastest firing MUAP and number MUAPs observed >10.

   Increased- Ratio of fastest firing MUAP and number MUAPs observed < 3.

4. MUAP Morphology Analysis:

Waveform morphology was observed with the needle electrode in close proximity to MUAP as evidenced by crisp sound and MUAP rise time <500us.

Normal- MUAPs consisting of 2-5 phases of #10 - 15ms in duration and <6000uV amplitude

Abnormal- Multiple MUAPs present with 6 or more phases > 10 - 15ms duration, Multiple MUAPs with amplitude > 6000uV, a combination of the previous two finding, or most MUAP's with amplitude <1000uV amplitude.

Classification: The results of the EMG/NCS examination consistent with CR and consistent with CTS were used to classify patients according to severity of findings for each respective condition.

1. Normal- No abnormalities noted
2. Unilateral median nerve (CTS) abnormalities:[127]

Mild- any abnormal median sensory latency or disto-proximal ratio. All other sensory and motor NCS parameters normal.

Moderate - abnormal sensory or disto-proximal ratio and distal motor latency. CNAP amplitude may be diminished but >50% of normal. Motor NCV normal.

Pronounced- abnormal sensory and distal motor latency. CNAP amplitude <50% of normal. CMAP amplitude may be diminished but >50% of normal. Mild slowing of forearm NCV may be present (>45 M/s) and spontaneous activity may be noted on EMG exam.

Severe- Absent CNAP, abnormal distal motor latency, CMAP amplitude <50% of normal or absent. Mild (>45 M/s) slowing of forearm NCV may be present and EMG abnormalities are present.

3. Bilateral median (CTS) abnormalities (each hand classified as above according to severity).
4.* Classification number 2 or 3 above with concomitant ulnar nerve abnormalities
5. Radiculopathy abnormalities:

Mild- H-reflex abnormality alone and/or 1+ spontaneous activity in one or more muscles. Other EMG/NCS parameters normal

Moderate- 2+ - 3+ spontaneous activity in two or more muscles. Increased recruitment, polyphasicity, and increased amplitude/duration of some MUAP's may be observed

Severe- 3+ - 4+ spontaneous activity in two or more muscles. Either increased recruitment ratio, polyphasicity, or increased amplitude/duration of many MUAP's is observed.

6.** Radiculopathy with concomitant CTS (double crush; both conditions classified according to their respective severity scales)

7.** Other: EMG/NCS studies consistent with other peripheral neuropathy or myopathy.

*Subjects classified in groups 4 & 7 will be eliminated from the study based on inclusion/exclusion criteria.

**Data from subjects classified in group 6 will be analyzed separately in a descriptive fashion.

4.2.4 Diagnostic Tests

4.2.4.1 Clinical Examination Procedures

Questions of History: All patients were asked 11 questions of history related to their signs and symptoms in the manner previously described. These 11 questions along with their respective possible responses are listed in appendix C.

Conventional Neurological Examination of the Upper Extremity:

Strength testing was conducted through manual muscle testing of the deltoid (C5), biceps brachii and extensor carpi radialis longus/brevis (C6), triceps brachii and flexor carpi radialis (C7), abductor pollicus brevis (C8), and dorsal interossei (T1). All manual muscle testing was conducted using the methods of Kendall and McCreary and performed with the subject sitting.[251] The deltoid was tested by resisting shoulder abduction. The biceps brachii was tested by resisting elbow flexion with the forearm supinated. The extensor carpi radialis longus/brevis was tested by resisting wrist extension from a neutral forearm position and 90^0 elbow flexion. The flexor carpi radialis was tested by resisting wrist flexion from a neutral forearm position and 90^0

elbow flexion. The triceps brachii was tested by resisting elbow extension from a position of 90^0 of elbow flexion. The abductor pollicus was tested with the forearm supinated and wrist in a neutral position. The first dorsal interossei was tested by resisting abduction of the first finger. The result of each muscle test was graded as "absent", "markedly reduced", "reduced", or "normal"

Muscle stretch reflexes of the bicep (C5-6), brachioradialis (C5-6), and Triceps (C7) were tested bilaterally using a standard reflex hammer. The result of each muscle stretch reflex was graded as "absent", "reduced", "normal", or "hyper/increased" as compared to the unaffected extremity .

Sensation testing was performed by testing sensitivity to light touch for the different cervical dermatomes (C5-C8) and discrete areas of median nerve cutaneous distribution (palmar surface of digits 1 – 3). Testing was performed by having the examiner touch the skin in a key area for each respective sensory level with a disposable paper clip that was discarded following testing. A new paper clip was used for each patient. Each dermatome level of the right and left upper limb was tested sequentially. The C5 dermatome was tested over the deltoid muscle, C6 along the radial aspect of the second metacarpal and index finger, C7 on the mid-posterior forearm and dorsal aspect of the middle finger, and C8 along the medial border of the 5^{th} finger. The discrete areas of median nerve cutaneous distribution were tested by comparing the palmar cutaneous distribution of digits 1-3 with the cutaneous distribution of the thenar eminence and midpalm area. The result of each sensory test was graded as "absent", "reduced", "normal", or "hyperesthestic" in comparison to the unaffected extremity.

Range-of-Motion and Wrist Diameter Measurements: The cervical ROM measures obtained in this study include: flexion, extension, bilateral side-bending, and bilateral rotation and were obtained in the following manner: While seated in a chair and prior to measurement by a physical therapist rater, the patient was asked to assume a neutral neck position satisfactory to both the patient and examiner. Once an acceptable neutral position has been assumed, the therapist applied a piece of colored tape to the wall at eye

level. This was referred to by the therapist as the "neutral position". The patient was then asked to perform warm-up movements consisting of two repetitions in each motion direction. Immediately following the warm-up procedure, the rater recorded a single ROM measurement for flexion, extension, and bilateral side-bending using a bubble inclinometer as described by Hole.[170] Bilateral rotation was measured using a standard long-arm goniometer

The wrist-ratio index is a proportion derived from the ratio of anterior-posterior (numerator) and medial-lateral (denominator) wrist width measured in centimeters. A single pair of sliding calipers was used to measure both anterior-posterior and medial-lateral wrist width. From these measurements a wrist ratio index was computed.

Provocation Tests: Provocative tests were performed sequentially according to operational definition. The starting order for testing was varied in a systematic fashion to prevent the confounding influence of order effects. Starting with the first subject, the Rater began the clinical examination by administering the first test or measure on the testing list. For the second subject, the rater began with the second clinical examination measure on the testing list. This procedure was continued for each successive subject.

An increase or decrease in symptoms refered to the symptoms associated with the patient's condition, not discomfort or pain associated with the test procedure that is unrelated to the patient's condition. The following phrase was used when the patient was questioned regarding the influence of a test procedure on their symptoms: "Did that increase or decrease your symptoms in any way?"

The following provocative tests were performed in this study:

1. Neck Compression Test (method A)
2. Neck Compression Test (method B)
3. Distraction Test
4. Shoulder Abduction Test
5. Valsalva Maneuver
6. Upper Limb Tension Test (ULLT A)

7. Upper Limb Tension Test (ULLT B)
8. Tinel's Sign (method A)
9 Tinel's Sign (method B)
10. Phalen's Test
11. Carpal Compression Test

4.2.4.2 Patient Self-Report Measures: Prior to the EMG/NCS examination, patients completed the following self-report measures: NDI, Brigham and Women's Hospital SSS/FSS, Hand diagram and 10cm VAS; and FABQ.

4.2.5 Patient Outcome Gold Standard

At six-weeks, a follow-up form was mailed to all patients. In addition to the GRCS, the form listed questions and corresponding responses inquiring about the patient's surgical and conservative treatment history since enrollment in the study.

All self-report forms, including the follow-up form, are listed in appendix B

4.3 Data Analysis

In addition to the analyses of diagnostic accuracy and predictive validity described below, descriptive statistics (mean, frequency, range, and standard deviation) were computed for all variables of the clinical examination, patient self-report instruments, and EMG/NCS findings, dependent upon the appropriate scale of measurement.

4.3.1 Hypothesis #1 - Reliability

The first hypothesis to be tested is the inter-rater reliability of all clinical examination items. Reliability coefficients for the patients self-report measures used in this study have been previously reported and will not be reassessed.

Reliability has been defined as the consistency of a measurement when all conditions are thought to be held constant.[252] Reliability reflects the degree to which a score is free from errors of measurement and may be described as the percentage of score that is information (signal) as opposed to random error (noise).[253] A reliable test or measure has at least three aspects: 1. Repeated measurement should be expected to repeat the same score on two different occasions; 2. Measures obtained can be depended on to give a

close approximation of the true score; 3. Allows one to generalize what will occur on future measurement occasions.[254] Reliability is a prerequisite for validity.[252] Therefore, validity can only be meaningfully interpreted when a measure is reliable.

The use of an unreliable measure may result in several undesirable consequences. Unreliable measures will attenuate correlations between variables and thereby diminish the ability to detect a relationship if one exists. A direct result of this attenuation is the need for increased sample sizes to obtain a significant effect in clinical trials. Unreliable measures will also contribute to biased samples.[255] Strube has described a number of different reasons that may cause a test or measure to be unreliable. Sources of unreliability include: examiners perform the test or measure differently; examiners perform the test or measure similarly but different standards are used as anchor points; and examiners enter data differently, resulting in coding errors.[256] Miscommunication and lack of understanding may also contribute to unreliability.

Another reason for unreliability is the lack of variability in the item of interest. Reliability indices (rxx) are a ratio of the variance of interest over the sum of the variance of interest plus error as depicted below.[254] It is clear from this ratio that in order for an index of reliability to be interpretable, there must first be variability to explain.

$$rxx = \frac{\text{true score variablity}}{\text{true score variability + error variablility}}$$

Finally, unreliability may result from disagreement among the raters. In this latter case, there may be no way to modify the procedure in order to achieve reliability and the measure will no longer be useful.[254]

Reliability estimates rely on measurement models and their assumptions. The measurement scale of the data determine which model is appropriate for obtaining a reliability coefficient.[254] Bartko has described three approaches for estimation of the reliability of nominal or categorical data.[257] The first is descriptive and merely computes

the percentage of agreement between raters. A second approach is to use a coefficient of association such as the Phi or Spearman Rho statistic for dichotomies or rank-order correlation.[243pp.446-449] The problems with the percent agreement approach are a lack of a standardized range for interpretation and no correction for chance agreements.[242] The correlation approach is also problematic: it only indicates the degree of association for paired scores, not agreement. The covariance of paired ratings may be very different than actual agreement if systematic error is present.[258] The third and recommended approach is the use of Cohen's Kappa statistic:[254,259] Kappa is interpreted as an intraclass correlation coefficient and represents the proportion of agreement among raters after chance agreement has been removed.[242] Kappa is expressed symbolically as:[242]

$$K = \frac{Po - Pc}{1 - Pc}$$

Where Po equals the observed proportion of agreement and Pc is the proportion of expected agreement based on chance; chance agreement increases as the variability of observed ratings decreases. Kappa values theoretically range from –1 to +1 but extreme values are often restricted by reduced variability of the data.[242,254] Positive Kappa values are interpreted as actual agreement beyond that expected by chance; values approximating zero indicate chance agreement and values less than zero are interpreted as agreement that is worse than that expected by chance.[242] Landis and Koch have proposed the following ranges of Kappa coefficients and corresponding strength of agreement associated with them:[260]

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Although the ranges they proposed are arbitrary, they have found acceptance in the measurement literature[242.254] and allow a consistent nomenclature for describing the strength of agreement associated with Kappa statistics.[260] Fleiss has simplified this descriptive scale in the following manner: values below 0.40 to represent poor agreement; values between 0.40 and 0.75 represent fair to good agreement beyond chance, and values greater than .075 represent excellent agreement beyond chance.[242] Some extended uses of Kappa include: allowance for more than two raters, different raters for each subject, and allowance for missing data.[135]

Two models of reliability have been described for interval and ratio scaled data.[247] Both models reflect the ratio of the variance of interest over the sum of the variance of interest plus error as previously described and produce a reliability statistic called an intraclass correlation coefficient (ICC), which is designated by the symbol R. The ICC provides a meaningful index of how dependably a measure maps onto or is correlated with the underlying characteristic being assessed.[254] An ICC is directly interpretable as a proportion of explained variance and describes the ability of a measure to differentiate between subjects.[255] The first model of reliability considered is the classic or psychometric theory of reliability in which every test score is considered to be composed of two parts: true score and error score. The error score is comprised of true random error and error from other sources. The psychometric theory of reliability treats all sources of error the same and makes no distinction between them. Alternatively, the generalizability theory of reliability encompasses a second model of reliability that allows the error score to be partitioned in to several sources of variability termed "facets".[254] A random effects ANOVA is utilized to partition the total variation in scores into separate components corresponding to the variables in the design. In this manner, error sources that exert a systematic influence can be estimated and separated from their error component.[254] All reliability estimates for interval and ordinal level measures in this study will be based on the generalizability theory of reliabilty. Similar to Landis and Koch and Fleiss, Portney and Watkins have described the following ranges of ICC and the strength of agreement associated with them: R <.75= Poor to Moderate; R .75 - .90=

Good; and R >.90= Excellent. For most clinical measurements R should exceed .90.[243p.514]

A measure that is highly reliable does not necessarily mean it is of value when applied to individual patients. In addition to acceptable reliability, the *accuracy* or *precision* of a measure is important. Accuracy or precision are synonymous terms related to reliability and refer to the variability of one person's score over measurement occasions.[254] Precision may be expressed as the standard error of measure (SEM) and is depicted symbolically as SEM = SDx (1-R)½, where SDx is the standard deviation of the measure of interest and R is the ICC, or reliability coefficient, of the measure of interest.[254] Estimates of both precision and reliability are important. Indeed, low reliability may be of little concern if the index of variability suggests the inconsistency of measurements occur in a relatively small range. Measurement methods should provide data that are both sufficiently reliable and precise.[261]

Reliability coefficients have been reported for only a few of the clinical examination procedures in this study. Therefore, estimates of reliability were obtained for all clinical examination procedures assessed by this study. The reliability coefficients for all clinical examination measures, with the exception of cervical range-of-motion and wrist ratio, were reported as a Kappa statistic. Kappa was also reported for ordered responses such as the CCT, Phalens test, the ULTT's, and selected questions of history in addition to Kappa for collapsed categories (i.e. dichotomy). The qualitative interpretation for Kappa described by Fleiss et. al. was used in this study. Measures with Kappa values above .75 were considered excellent or exceptionally reliable; those with Kappa values between 0.40 and 0.75 were considered to have fair to good reliability; and those below 0.40 were considered poor and unreliable.[242] The reliability coefficient for cervical range-of-motion and the wrist ratio was reported as an ICC (2,1) statistic along with the corresponding SEM for both measures.[262] The reliability of cervical range-of-motion and wrist ratio measurements were considered excellent and well suited for clinical use if an ICC of .90 is achieved.[243p.514]

Point estimates of the reliability coefficient were used to decide whether to accept or reject research hypotheses of reliability. However, the confidence interval method was used to determine whether the point estimate of a reliability coefficient was a definitive finding. A definitive finding is considered to be a finding or value which would within a given range 95% of the time with repeated sampling.[243pp 292 - 294] Nintey-five percent CI's were computed for all reliability coefficients. When the lower bound of the 95CI for a given variable was equal to or above the upper limit of the hypothesized level of reliability for that variable, then the point estimate for that level of reliability was considered to be definitive. When the upper bound of the 95CI for a given variable was equal to or below the lower limit of the hypothesized level of reliability for that variable, then the point estimate for that level of reliability was also considered definitive.

4.3.2 Hypothesis #2 - Diagnostic Accuracy

The second hypothesis to be tested relates to the diagnostic properties of the clinical examination items and patient self-report instruments. The test items and condition they are intended to diagnose or evaluate are listed on the following page.

Cervical Radiculopathy

Questions

1.- "Most bothersome symptoms.."
2.- "Where most bothersome.."
3.- "Symptom behavior.."
6.- "Entire limb numb.."
7.- "Symptoms keep from sleep.."
9..- "Neck movement improves.."

Neurologic exam: sensory (dermatomes), motor, and reflexes

Measures: ROM (flexion, extension, sidebending, and rotation)

Provocation tests: Spurling's A & B, Distraction, Shoulder abduction, Valsalva, ULTT A & B

HSAM/Self-report Instruments: NDI, 10cm VAS, FABQ

Carpal Tunnel Syndrome

Questions:

1.- "Most bothersome symptoms.."
2.- "Where most bothersome.."
3.- "Symptom behavior.."
4.- "Hand fat/swollen.."
5.- "Fumbling/dropping.."
6.- "Entire limb numb.."
8.- "Night symptoms wake.."
10.- "Hand shaking improves.."
11.- "Worse with hand use.."

Neurologic exam: sensory and (dermatomes and median distribution), motor

Measures: Wrist ratio

Provocation tests: ULTT A & B, Tinel's A & B, Phalen's, CCT

HSAM/Self-report Instruments: SSS & FSS, Hand diagram, 10cm VAS, FABQ

The diagnostic properties of a single cluster of test items (test item cluster (TIC)) identified by logistic regression as being the best predictor of each respective condition was also evaluated. Additional variables of age and duration of symptoms were included in this analysis

Sensitivity and Specificity: Two-by-two contingency tables were used to calculate Sn and Sp for each test item relative to their respective condition, either CR or CTS, for the total sample. Ninety-five percent confidence intervals were also calculated for all Sn and Sp values. In this study, the presence of each condition as determined by EMG/NCS findings constituted the gold standard to which the positive and negative findings of each test item were compared. For example, all patients classified as CTS were considered to have the condition present and the remainder of patients the condition was considered absent. Because diagnostic tests may have different sensitivities or specificities in different parts of the clinical spectrum of the disease they purport to identify or exclude,[50] Sn and Sp were also be calculated for subgroups of patients in both conditions based on severity of EMG/NCS findings. Patients classified as mild or moderate CR or CTS formed one subgroup within each respective condition. Patients classified with pronounced or severe CTS and patients classified as severe CR formed the other subgroup. Subgroup calculations were only be performed if the prevalence of condition was 10% or greater for a given subgroup.[24pp.90-91] Sensitivity and specificity calculations have been previously described in section 2.2.3.

To avoid confounding, all patients determined to have both CR and CTS were excluded from the diagnostic accuracy analysis and were reported as a percentage or frequency statistic of the total sample. The concomitant presence of both conditions has been described in the literature and is often referred to as the "double-crush" phenomenon.[18,20,263,264] Based on previous reports, the percentage of patients in this study expected to have CR and CTS concomitantly is approximately 3-5%.[20,263]

Receiver Operating Characteristic (ROC) Curve: In order to identify the most appropriate cut-off value for continuous or multi-level response variables, the Sn and Sp

values for each level of response were plotted as a ROC curve. The ROC curve plots sensitivity (true positive ratio) against 1 – specificity (false positive ratio) for the criterion defining a positive test.[265] An ROC curve is simply a graph of the pairs of true positive rates and false positive rates that correspond to each possible cutoff value for the diagnostic test result.[24pp.117] The area may range from .5 (no diagnostic ability) to 1.0 (perfect diagnostic ability) as the ROC curve moves towards the top left-hand corner of the graph;[266] the area under a ROC curve represents the diagnostic ability of the test. Receiver operating characteristic curves were constructed for the following variables: Cervical range-of-motion, wrist ratio, and all patient self-report measures. The value closest to the upper left-hand corner of the graph minimizes the occurrence of false positive and false negatives when the prevalence rate is around 50%.[24pp.118]

Likelihood ratios (LR+ & LR-) were calculated for each test item in the manner previously described in section 2.2.3 along with their associated 95CI.

TIC Cluster: Clinical diagnosis rarely resides in a single finding, but more often in the pattern of findings.[27] Therefore, using a combination of tests as a single TIC may increase the diagnostic value of the tests.[16,267p.41] Because each component of the clinical examination can be considered a separate test, one must choose how to incorporate the numerous results.[23] One method of combining various items of the clinical examination is the use of LR's to sequentially modify posterior probability of the presence or absence of a target disorder and was illustrated in the example in section 2.2.3. However, this serial multiplication of LR's assumes that the tests are conditionally independent.[24p.133] Conditional independence means the result of one test is not affected by the outcome of any of the other tests performed.[23] If the assumption of test independence is violated, diagnostic accuracy can be degraded and result in inaccurate assessments.[268,24pp136-139]

The method described by Holleman and Simel was used to identify the most accurate TIC's: one TIC for the diagnosis of CR and another for the diagnosis CTS. To reduce the number of variables, LR's for test items with a 95%CI that included .60 to 1.4 were be excluded from further consideration; LR values of one or close to one are indeterminate

and therefore are not considered useful.[23] Remaining variables were then entered using a backward stepwise procedure into a binary logistic regression model (condition present or absent). Variables selected by the regression model as most predictive of the condition of interest were combined or clustered into a TIC and treated as a single test item. The sensitivity, specificity, and LR's for the TIC was computed as previously described. Although the variables identified by this method may still be interdependent to some degree, Holleman and Simel reported no difference in prediction ability between this method and a more complex procedure that forced conformity with the independence assumption.[23]

In this study, point estimates and 95% CI's for the Sn, Sp, likelihood ratios (LR+, LR-), were computed for each clinical examination item, patient self report measure, and TIC.

The diagnostic accuracy values of individual clinical examination variables were considered acceptable for their respective condition when any of the following occured: 1. Either Sn or Sp is equal to or greater than .70; 2. A LR+ equal or greater than 2.0; and 3. A LR- equal to or less than .50. When test Sn and Sp values both equal or exceed .70, LR+ and LR- values will exceed 2.3 and be below .43, respectively. Based on the estimated prevalence or pretest probability of CR and CTS in this sample, LR+ values >2.0 and LR- values <.5 will result in posttest probability changes of at least 15%.

The guidelines listed in the preceding paragraph were used to accept or reject the previously specified hypothesis of diagnostic accuracy for an individual clinical examination variable as well as determine whether a diagnostic accuracy value is considered defninitve.

4.3.3 Hypothesis #3 - Test Predictive Validity

The third hypothesis to be tested relates to the predictive validity of the clinical examination items and patient self-report instruments. In addition, the FABQ score and the following EMG/NCV variables were also included as predictor variables: sensory and motor nerve conduction latency; sensory and motor amplitudes; and presence of

spontaneous activity in the abductor pollicus brevis. The diagnostic properties of a single TIC identified by logistic regression as being the best predictor of patient outcome was also evaluated.

Although current gold standards tend to be defined in terms of pathologic anatomy, clinical course and prognosis of patients are increasingly used as gold standards.[24pp.81] Predictive validity for each test item and a single TIC were reported as Sn, Sp, and LR's. In this study, the following analyses were performed for both the CR group and the CTS group to establish the predictive validity of the pertinent test items.

For the first analysis, type of intervention, defined as surgical or non-surgical, served as the gold standard. Patients who received surgery were considered positive and those treated non-surgically were considered negative. If the prevalence of surgery for either condition is less than 10%, an analysis of predictive validity using type of intervention was not be performed for that condition. The ability of a test to produce a meaningful change in posttest probability for a condition is severely diminished when prevalence of the condition is at either extreme.[24pp.90-91]

For the second analysis, patient outcome defined by patient improvement using a GRCS was the gold standard. The criteria recommended by Jaeschke et. al. was used to determine subject improvement: subjects scoring between –5 and +3 ("somewhat worse" and "somewhat better", respectively) were considered unimproved (stable, no meaningful change in condition) or to have deteriorated. Subjects scoring greater than +3 ("moderately better" to "a very great deal better") were considered improved or to have undergone clinically meaningful change. Patients who were unimproved or worsened were considered negative and those who are improved were considered positive. Two separate analyses of patient improvement were performed, one for non-surgically treated subjects and another or surgically treated subjects. The patient improvement analysis of surgically treated subjects was not be performed if surgery prevalence for either condition was less than 25 subjects or less than 10%.

For the third analysis, patient outcome defined by worsened patient condition using a GRCS served as the gold standard. The criteria recommended by Jaeschke et. al. was used to determine subject improvement: subjects scoring between –5 and +3 ("somewhat worse" and "somewhat better", respectively) were considered unimproved (stable, no meaningful change in condition). Subjects scoring less than -5 ("moderately worse" to "a very great deal worse") were considered worsened or to have undergone clinically meaningful change. Patients who are worsened were considered positive and those who are unchanged or improved were considered negative. Two separate analyses of worsened patient condition were performed, one for non-surgically treated subjects and another or surgically treated subjects. The patient improvement analysis of surgically treated subjects was be performed if surgery prevalence for either condition was less than 25 subjects or less than 10%.

Two-by-two contingency tables for Sn and Sp, ROC curves, Likelihood ratios (LR+; LR-), and identification of a single test item cluster were computed in the manner previously described for hypothesis #2. The same criteria used for hypothesis #2 was used to accept or reject the previously specified hypothesis of predivitive validity for an individual clinical examination variable or the TIC and to determine which variables were to be considered definitive.

All statistical test procedures were computed using Microsoft Excel and the SPSS statistical software packages.

5.0 SAMPLE SIZE ESTIMATION

Because this study is descriptive in nature, a sample size estimate derived from power calculations based on group differences is not possible. Instead, sample size was based on the ability of this study to detect the following: 1. An ICC of .90 that is significantly greater than .75 at an alpha level of .05 and beta level of .20 for a one-tailed test (i.e. power is greater than .80);[243p.514] 2. A Kappa coefficient of .60 that is significantly greater than .40 at an alpha level of .05 and beta level of .20 for a one-tailed test with a base chance-agreement rate of .50 (i.e. power is greater than .80);[242] and 3. A test

sensitivity or specificity of .80 whose 95%CI has a minimum lower bound that exceeds .68. Forty subjects (20 of each condition) from each of the following facilities will be required for this study (160 total subjects): The National Naval Medical Center (NNMC), Bethesda, MD; Wilford Hall USAF Medical Center (WHMC), San Antonio, TX; and the Air Force Academy Hospital (AFAH), Colorado Springs, CO The procedure described by Kraemer and Thiemann and implemented by the EX-SAMPLE statistical computer package indicates that this sample size is more than adequate to establish the specified reliability coefficients for each facility.[269]

Based on the estimated prevalence rates for each condition in this study, a sample size of 160 is the minimum for which the lower bound of the 95% confidence interval for a true sensitivity or specificity of .80 would exceed .68.[55]

6.0 RESULTS

6.1 Study Sample and Diagnostic classification

A total of 81 patients from the following three participating medical centers met eligibility criteria and were enrolled in the study: Wilford Hall Medical Center (n= 68), EMG laboratories of both Montefiore and Presbyterian University Hospital (n= 11), and Brooke Army Medical Center (BAMC) (n=2). Two of the original participating centers, the National Naval Medical Center and the United States Air Force Academy Hospital, eliminated themselves during the course of the study due to subject enrollment difficulties. One additional facility, BAMC, participated in subject enrollment after the study commenced. Due to limited subject enrollment at all facilities, the original study entry criteria for duration of symptoms was eliminated and duration of symptoms was recorded for all subjects. The Institutional Review Board of all participating facilities approved all changes to the study protocol. All consecutive patients referred to the EMG lab with suspected CR, CTS, or with other suspected diagnosis but had symptoms compatible with CR or CTS were informed about the study by EMG lab personnel. Interested subjects were asked to fill out a screening form to determine eligibility (appendix A). Interested and eligible subjects were given further information about the

study, then read and signed an informed consent document approved by the respective facility Institutional Review Board.

The frequency of conditions suspected by the referring provider is compared with the conditions suspected by the EMG/NCS provider in Table 6; the EMG/NCS provider suspected condition was not available for 5 subjects. Although the same number of subjects were suspected by the referring providers to have CR, there was not always concordance between the two providers. The referring provider suspected CTS in three subjects diagnosed with CR while the EMG/NCS provider suspected CR, normal, and both conditions for these same individuals. None of the subjects who participated in this study were receiving workman's compensation or had pending litigation for their condition.

Table 6. Condition suspected by providers

	Referring Provider	Electromyography Provider
Condition	Frequency	
Radiculopathy	29	29
CTS	42	31
Both	5	7
Other	5	9
Total Available	81	76

Seven different EMG/NCS providers performed the nerve conduction studies, needle electromyography procedures, and subsequent diagnostic classification of subjects. At one center, three different evoked potential technicians performed nerve conduction procedures only. The qualifications of the EMG/NCS providers and evoked potential technicians are listed in Table 7. All EMG/NCS diagnostic classifications made by non-physician EMG/NCS providers were reviewed and approved by the supervising EMG/NCS lab physician who is board certified by the American Academy of Electrodiagnostic Medicine (AAEM). There were no disagreements between the non-physician EMG/NCS providers and supervising EMG/NCS lab physicians regarding diagnostic classification of subjects.

Table 7. Qualifications of personnel performing electrophysiologic testing

Facility and Provider Number, and Role	Years EMG/NCS Experience	Board Certification	Number of studies performed
University of Pittsburgh			
1. EMG	5	Yes: AAEM	4
2. EMG	7	Yes: AAEM	2
3. EMG	2	Yes: AAEM	3
4. EMG	6	Yes: AAEM	1
5. EMG	4	Yes: AAEM	1
6. NCS.	15	Yes: AAEM	6
7. NCS	5	Yes: AAEM	4
8. NCS	2	Yes: AAEM	1
Wilford Hall Medical Center			
1. EMG/NCS	10	Yes: ABPTS ECS	49
2. EMG/NCS	17	Yes: ABPTS ECS	19
Brooke Army Medical Cetner			
1. EMG/NCS	17	Yes: ABPTS ECS	2

AAEM: American Academy of Electrodiagnostic Medicine
ABPTS ECS: American Board of Physical Therapy Specialties, Electrophysiologic Certified Specialist

Forty-one females (mean age= 44.9yrs, sd= 12.5 range= 24 –70) and 40 males (mean age= 45.0yrs, sd=11.4, range= 21 – 68) participated in this study. Once enrolled in the study, subjects completed all self-report instruments and received a standardized EMG/NCS examination. Following the standardized EMG/NCS examination, subjects were assigned to the following diagnostic categories based on the results of the EMG/NCS examination and the assessment/impression of the EMG/NCS provider: 1. Normal (n= 31), 2. Unilateral CTS (n= 16), 3. Bilateral CTS (n= 15), 4. CTS with ulnar neuropathy (n= 1), 5. Cervical radiculopathy (n= 13), 6. Cervical radiculopathy with CTS (n= 3), 7. Other (n= 2).

The subjects age, duration of symptoms, and several nerve conduction study parameters of the median nerve are compared in Tables 8 and 9 by diagnostic category and gender.

Table 8. Descriptive statistics of subjects age and duration of symptoms.

EMG/NCS based Dx	Gender	N	Variable Age= years Symptoms= days	Mean Median	Minimum	Maximum	Sd Dev.
Normal	Female	17	Age	39.29	24.00	61.00	12.14
			Symptoms	123.5	31.00	5415.00	
	Male	14	Age	38.78	21.00	68.00	10.94
			Symptoms	184.5	21.00	7220.00	
Unilateral CTS	Female	10	Age	51.90	31.00	70.00	12.97
			Symptoms	352	21.00	1460.00	
	Male	6	Age	38.00	28.00	49.00	7.89
			Symptoms	365	56.00	1277.00	
Bilateral CTS	Female	9	Age	44.77	28.00	61.00	11.53
			Symptoms	250	31.00	5475.00	
	Male	6	Age	47.16	36.00	60.00	10.00
			Symptoms	61	21.00	365.00	
CTS w/ Ulnar neuropathy	Female	1	Age	43.00	43.00	43.00	.
			Symptoms	30.00	30.00	30.00	.
Radiculopathy	Female	2	Age	56.50	55.00	58.00	2.12
			Symptoms	42.00	42.00	42.00	
	Male	11	Age	50.90	39.00	61.00	7.68
			Symptoms	69.5	42.00	1095.00	
Radiculopathy w/CTS	Female	1	Age	52.00	52.00	52.00	.
			Symptoms	.	.	.	
	Male	2	Age	62.00	60.00	64.00	2.82
			Symptoms	31.50	21.00	42.00	
Other	Female	1	Age	46.00	46.00	46.00	.
			Symptoms	87.00	87.00	87.00	.
	Male	1	Age	62.00	62.00	62.00	.
			Symptoms	551.00	551.00	551.00	.

Table 9. Descriptive statistics of subjects median nerve conduction study test results.

EMG/NCS based Dx	Median Nerve Parameters*	N	Mean	Minimum	Maximum	Std. Deviation
Normal	Involved palmar latency	31	1.89	1.60	2.20	.15
	Involved motor latency	31	3.62	2.90	4.30	.41
	Involved palmar amplitude	31	81.09	40.00	140.00	25.80
	Involved motor amplitude	31	10229.35	5000.00	18650.00	3139.99
	Distal-proximal ratio	27	.78	.60	.90	.10
Unilateral CTS	Involved palmar latency	16	3.28	1.80	10.00	2.65
	Involved motor latency	16	4.95	3.30	14.90	2.78
	Involved palmar amplitude	16	62.62	.00	183.00	46.34
	Involved motor amplitude	16	7106.25	100.00	10770.00	2937.15
	Distal-proximal ratio	12	.99	.60	1.50	.23
Bilateral CTS	Involved palmar latency	15	3.68	2.30	10.00	2.59
	Involved motor latency	15	4.79	4.00	6.40	.77
	Involved palmar amplitude	15	46.53	.00	114.00	33.63
	Involved motor amplitude	15	8446.00	4900.00	13300.00	2522.23
	Distal-proximal ratio	11	1.21	1.00	1.60	.21
CTS w/ Ulnar neuropathy	Involved palmar latency	1	2.30			.
	Involved motor latency	1	3.20			.
	Involved palmar amplitude	1	42.00			.
	Involved motor amplitude	1	10000.00			.
	Distal-proximal ratio	1	1.30	1.30	1.30	.
Radiculopathy	Involved palmar latency	13	1.99	1.80	2.20	.16
	Involved motor latency	13	3.73	3.20	4.20	.35
	Involved palmar amplitude	13	72.92	40.00	160.00	41.46
	Involved motor amplitude	13	10495.38	5780.00	18240.00	3654.60
	Distal-proximal ratio	13	.76	.60	.90	.10

Table 9 (cont'd)

Radiculopathy w/CTS	Involved palmar latency	3	3.03	2.80	3.20	.2082
	Involved motor latency	3	4.46	4.40	4.50	5.77
	Involved palmar amplitude	3	43.66	38.00	52.00	7.3711
	Involved motor amplitude	3	10500.00	8490.00	12000.00	1809.7237
	Distal-proximal ratio	1	1.30			
Other	Involved palmar latency	2	6.25	2.50	10.00	.01
	Involved motor latency	2	6.55	4.80	8.30	2.47
	Involved palmar amplitude	2	18.00	.00	36.00	25.45
	Involved motor amplitude	2	7230.00	4720.00	9740.00	3549.67
	Distal-proximal ratio	2	.70	.00	1.40	.98

***Latency in milliseconds and amplitudes in microvolts.**

Thirteen subjects (16%) were classified with CR, the left extremity was involved in nine subjects and the right in three. These CR subjects were suspected by the EMG provider to have the following conditions prior to the EMG/NCS examination: CR= 10, CTS= 1, both conditions= 1, other= 1. The following conditions were suspected for these same 13 subjects by the provider who referred the patient to the EMG lab: CR= 9, CTS= 4. The 13 subjects diagnosed with CR and 31 subjects diagnosed with CTS were further subclassified based on the severity of EMG/NCS findings. For the 13 CR subjects, nine were subclassified as mild and four as moderate. The frequency of needle EMG findings for muscles tested in the standardized exam are listed in Table 10. Needle EMG testing of muscles other than those specified in the standardized EMG/NCS exam was permitted when thought indicated by the EMG/NCS provider. The additional muscles sampled, along with frequency and findings, and are listed in Table 11. In only one instance did additional muscle testing yield abnormal findings (brachioradialis) when the results from the previously tested standardized muscles was normal. Two subjects were unable to

tolerate EMG sampling of the middle cervical paraspinal muscles and one subject was unable to tolerate EMG sampling of the lower cervical paraspinal muscles. Because the flexor carpi radialis H-reflex was technically unobtainable in 42% of the subjects, it was eliminated as part of the diagnostic criteria for CR. The diagnostic report for all 13 subjects indicated involvement of the C6 or C7 root, possible involvement of the C8 root in 1 subject, and C5 root in two subjects.

Thirty-one subjects were classified as CTS, the left extremity was involved in 11 subjects and the right in 20. These CTS subjects were suspected by the EMG provider to have the following conditions prior to the EMG/NCS examination: CTS= 19, CR= 4, both= 3, and other=3; ratings for two subjects were missing. The following conditions were suspected for these same 31 subjects by the referring provider: CTS= 20, CR= 6, both= 3, and other=2. Of the 31 CTS subjects, 14 were classified as mild, 7 as moderate, 9 as pronounced, and 1 as severe. The subclassification of CTS subjects is presented in Tables 12 and 13 along with descriptive statistics for several electrophysiologic parameters, age, and duration of symptoms. Needle EMG findings in the adbuctor pollicus brevis are listed by subclassification in Table 14. There were five CTS subjects whose only abnormal NCS finding was a > .2ms median/ulnar palmar latency difference, and four subjects whose only abnormal NCS finding was a distal-proximal ratio >1.0. The diagnosis in the remainder of these subjects was based on a prolonged median palmar latency and/or other concomitant abnormal median nerve conduction study parameters. All subjects diagnosed with bilateral CTS had symptoms that were predominate in one hand. The hand with predominate symptoms was considered to be the involved limb for the purposes of this study and was used for subclassification and subsequent clinical testing.

Following the standardized EMG/NCS examination and approximately a 15 to 30 minute rest period, a standardized clinical examination was performed by a physical therapist (Rater 1) and repeated by a second physical therapist (Rater 2) following a five to ten minute rest period. Both raters were blinded to the subjects suspected diagnosis, EMG/NCS test results, and diagnostic classification. Nine different physical therapists

performed the standardized clinical examination procedures. Distinct rater pairs were formed although substitution of raters did occur. The qualifications of the physical therapist raters are listed in Table 15.

Table 10 Frequency of needle electromyography findings in standardized examination muscles.

Muscle Tested	Silent at Rest	1+ Fibs/PSW	2+ Fibs/PSW	3+ Fibs/PSW	4+ Fibs/PSW
Deltoid	11	1	1		
Triceps brachii	11		2		
Extensor carpi radialis longus/brevis	12		1		
Flexor carpi radialus	10	2	1		
Abductor pollicus brevis	13				
First dorsal interosseus	13				
Middle cervical paravertebrals muscles	6	3	3		
Lower cervical paravertebrals	6	3	1	1	

Fibs/PSW- Fibrillation potentials and positive sharp waves

Table 11 Frequency of needle electromyography findings in muscles sampled in addition to the standard examination muscles.

Muscle Tested	Silent at Rest	1+ Fibs/PSW	2+ Fibs/PSW	3+ Fibs/PSW	4+ Fibs/PSW
Pronator Teres	4				
Biceps Brachii	6	1	1		
Supraspinatus	4				
Infraspinatus	4	2			
Flexor carpi ulnaris	3	1			
Extensor indicis proprius	1	1			
Flexor digitorum profundus (slips 3 & 4)	1				
Brachioradialis		1			

Fibs/PSW- Fibrillation potentials and positive sharp waves

Table 12 Descriptive statistics of median nerve conduction study test results by carpal tunnel syndrome subclassification category

CTS subclassification	Median Nerve Parameters*	N	Mean	Minimum	Maximum	Std. Deviation
Mild	Involved palmar latency	14	2.24	1.80	2.70	.27
	Involved motor latency	14	3.90	3.30	4.60	.32
	Involved palmar amplitude	14	75.00	40.00	183.00	42.07
	Involved motor amplitude	14	8419.28	5800.00	12600.00	2073.79
	Distal-proximal ratio	14	.95	.60	1.20	.15
Moderate	Involved palmar latency	7	2.75	2.10	3.40	.51
	Involved motor latency	7	4.75	4.10	5.40	.41
	Involved palmar amplitude	7	63.00	39.00	114.00	28.70
	Involved motor amplitude	7	8044.28	4900.00	13300.00	2991.82
	Distal-proximal ratio	5	1.22	1.20	1.30	.004
Pronounced	Involved palmar latency	9	5.23	2.40	10.00	3.59
	Involved motor latency	9	5.36	4.40	6.70	.89
	Involved palmar amplitude	9	23.22	.00	52.00	22.21
	Involved motor amplitude	9	7345.55	1900.00	10290.00	2642.84
	Distal-proximal ratio	4	1.47	1.20	1.60	.18
Severe	Involved palmar latency	1	absent	--	--	.
	Involved motor latency	1	--	14.90	14.90	.
	Involved palmar amplitude	1	.00	.00	.00	.
	Involved motor amplitude	1	100.00	100.00	100.00	.
	Distal-proximal ratio	0				

***Latency in milliseconds and amplitudes in microvolts.**

Table 13 Descriptive statistics of subjects age and symptom duration by subclassifcation

Carpal Tunnel subclassification	Variable Age= years Symptoms= days	N	Mean ------- Median	Minimum	Maximum	Std. Deviation
Mild	Age	14	47.5714	31.00	70.00	11.7651
	Symptom Duration		154.5	154	5475	
Moderate	Age	7	35.7143	28.00	46.00	6.2906
	Symptom Duration		184	184	2555	
Pronounced	Age	9	53.6667	42.00	68.00	9.0830
	Symptom Duration		155	21	1460	
Severe	Age	1	34.0000	34.00	34.00	
	Symptom Duration					

Table 14 Frequency of spontaneous activity in the abductor pollicus brevis of subjects with CTS

CTS subclassification	Spontanteous Activity	Frequency
Mild	Silent at rest	14
Moderate	Silent at rest	6
	1+ Fibs/PSW	1
Pronounced	Silent at rest	5
	1+ Fibs/PSW	3
	2+ Fibs/PSW	1
Severe	4+ Fibs/PSW	1

Table 15 Qualifications of physical therapist raters

Center number and Therapist Number	Years of Clinical Experience	Board Certification	Number First Examinations performed	Rater Pairs and Number of Paired Examinations Performed
University of Pittsburgh (1)				
1	7	No	4	1 & 3= 5
2	15	Yes: ABPTS OCS	7	1 & 5= 1
3	9	Yes: ABPTS OCS, SCS	0	
Brooke Army Medical Center (2)				
1		Yes: OCS		1
2	15	Rater 3, facility 1	2	
Wilford Hall Medical Center (3)				
1	24	Yes: ABPTS OCS	56	1 & 2= 3 1 & 3= 50
2	15	Yes: ABPTS OCS	1	1 & 4= 1 2 & 3= 1 3 & 4= 4
3	13	Yes: ABPTS OCS	6	
4	25	Yes: ABPTS OCS, SCS,	5	

ABPTS OCS: American Board of Physical Therapy Specialties, Orthopaedic Certified Specialist
ABPTS SCS: American Board of Physical Therapy Specialties, Sports Certified Specialist

6.2 Hypothesis #1 – Reliability

Rater pair 1 & 3 at Wilford Hall Medical Center performed 50 examinations together. results from rater pair 1 & 3 were used to compute all reliability statistics due to the

limited number of examinations performed by other specified rater pairs. The 31 subjects not included in the reliability analysis did not differ from the other 50 subjects with regard to age, NDI, SSS, FSS, or pain ratings (all p values >.05).

The reliability statistics for clinical examination variables are shown in Table 16 and include SEM values of continuous measures. The stem-and-leaf plots in Figures 15 – 19 show the distribution of scores for continous measures. Due to low observed base-rates in uninvolved subjects, only tests results of the involved limb of were used to compute reliability statistics for the neurologic clinical examination variables.[270] Due to asymmetric contingency tables, reliability statistics were only computed using collapsed categories (3 levels) and dichotomized results for ULTT A, ULTT B, and Phalen's test. Due to the low observed base-rates of increased sensation (rater 1 n=4, rater 3 n= 5) and hyper-reflexia (rater 1 n=2, rater 3 n= 0), results for dermatomes, median nerve fields, and MSR's were dichotomized into normal (normal or hyper) or abnormal (reduced) findings. There were no manual muscle test (MMT) scores of zero. Rater 1 identified 3 subjects with P- to F ratings and the remainder of subjects received ratings in the other two categories. Therefore, reliability statistics were computed for MMT scores that were dichotomized as normal and abnormal. No abnormal findings for the tricep and brachioradialis muscle stretch reflexes were recorded for rater pair 1 & 3 so reliability could not be computed for these variables. Reliability statistics were also computed for transformed ratings that identified an abnormality of any dermatome, myotome or median sensory field of the involved limb. The reliability of these variables was fair to good with Kappa values of .51, .64, and .48, respectively. Although reliability for MMT of the abductor pollicus brevis was poor when assessed in subjects with a variety of conditions, it demonstrated fair to good reliability when assessed in subjects diagnosed with CTS (Kappa= .65). However, the observed base rate was still low (valid n=22, observed base rate of 10%).

Table 16 Reliability of clinical examination variables with 95% confidence intervals (95CI)

Variable	Kappa 95CI	ICC 95CI	SEM
question 1-"Most bothersome symptoms.."	.74 (.55 - .93)	–	–
question 2-"Where most bothersome.."	.82 (.68 - .96)	–	–
question 3-"Symptom behavior.."	.57 (.35 - .79)	–	–
question 4-"Hand fat/swollen.."	.85 (.68 - 1.0)	–	–
question 5-"Fumbling/dropping.."	.95 (.85 - 1.0)	–	–
question 6-"Entire limb numb.."	.53 (.26 - .81)	–	–
question 7-"Symptoms keep from sleep.."	.70 (.48 - .92)	–	–
question 8-"Night symptoms wake.."	.83 (.60- 1.0)	–	–
question 9-"Neck movement improves.."	.67 (.44 - .90)	–	–
question 10-"Hand shaking improves.."	.90 (.75 - 1.0)	–	–
question 11-"Worse with hand use.."	.72 (.49 - .95)	–	–
C5 Dermatome	.67 (.33 - 1.0)	–	–
C6 Dermatome	.28 (.00 - .58)	–	–
C7 Dermatome	.40 (.06 - .74)	–	–
C8 Dermatome	.16 (.00 - .50)	–	–
T1 Dermatome	.46 (.04 - .88)	–	–
MMT deltoid	.62 (.28 - .96)	–	–
MMT biceps brachii	.69 (.36 - 1.0)	–	–
MMT extensor carpi radialis longus/brevis	.63 (.26 - 1.0)	–	–
MMT triceps brachii	.29 (.00 - .79)	–	–
MMT flexor carpi radialis	.23 (.00 - .69)	–	–
MMT abductor pollicus	.39 (.00 - .80)	–	–

Table 16 (cont'd.).

Variable	Kappa 95CI	ICC 95CI	SEM
MMT first dorsal interrosseus	.37 (.00 - .80)	–	–
biceps brachii MSR	.73 (.38 - 1.0)	–	–
median sensory field 1	.48 (.23 - .73)	–	–
median sensory field 2	.50 (.25 - .75)	–	–
median sensory field 3	.40 (.12 - .68)	–	–
dermatome (any abnormal)	.51 (.26 - .76)	–	–
MMT (any abnormal)	.64 (.35 - .93)	–	–
Median nerve field (any abnomal)	.48 (.22 - .74)	–	–
Spurling's A	.60 (.32 - .87)	–	–
Spurling's B	.62 (.25 - .99)	–	–
Shoulder abduction	.20 (.00 - .59)	–	–
Valsalva	.69 (.36 - 1.0)	–	–
Distraction	.88 (.64 - 1.0)	–	–
Tinel A	.47 (.21 - .72)	–	–
Tinel B	.35 (.10 - .60)	–	–
ULTT A (collapsed)	.70 (.51 - .89)	–	–
ULTT A (dichotomized)	.76 (.51 - 1.0)	–	–
ULTT B (collapsed)	.45 (.23 - .67)	–	–
ULTT B (dichotomized)	.83 (.65 - 1.0)	–	–
CCT	.68 (.59 - .86)	–	–
CCT (dichotomized)	.77 (.58 - .96)	–	–
Phalen's (collapsed)	.44 (.26 - .62)	–	–
Phalen's (dichotomized)	.79 (.59 - 1.0)	–	–
cervical flexion	–	.79 (.65 - .88)	4.6 degrees
cervical extension	–	.84 (.70 - .95)	4.8 degrees
Cervical left rotation	–	.75 (.59 - .85)	6.6 degrees
Cervical right rotation	–	.63 (.22 - .82)	7.3 degrees
Cervical left sidebending	–	.63 (.40 - .78)	5.3 degrees
Cervical right sidebending	–	.68 (.62 - .87)	5.4 degrees
Wrist anterior-posterior	–	.77 (.62 - .87)	2.1 millimeters
Wrist medial-lateral	–	.86 (.75 - .92)	2.1 millimeters

```
Frequency     Stem &  Leaf

     3.00       3 .  000
     3.00       3 .  555
     8.00       4 .  00000000
    14.00       4 .  55555555555558
    16.00       5 .  0000000000000123
    10.00       5 .  5555555568
    15.00       6 .  000000000011223
     6.00       6 .  555555
     4.00       7 .  0000
     1.00       7 .  5

  Stem width:     10.00
  Each leaf:       1 case(s)
```

Figure 12. Distribution of cervical flexion measurements

```
Frequency     Stem &  Leaf

     2.00 Extremes    (=<25)
     5.00       3 .  00558
     9.00       4 .  000055556
    24.00       5 .  000000000445555555555588
    24.00       6 .  000000000000445555555558
    15.00       7 .  000000000000055
     1.00       8 .  0

  Stem width:     10.00
  Each leaf:       1 case(s)
```

Figure 13. Distribution of cervical extension measurements

```
Frequency     Stem &  Leaf

     4.00       3 .  0055
    13.00       4 .  0000355555555
    21.00       5 .  000000000000222335557
    25.00       6 .  0000000000000055555555689
     9.00       7 .  000055568
     8.00       8 .  00000225

  Stem width:     10.00
  Each leaf:       1 case(s)
```

Figure 14. Distribution of cervical left rotation measurements

```
Frequency     Stem &  Leaf

     6.00       3 .  000557
    17.00       4 .  00000000555555778
    30.00       5 .  000000000000003455555555555567
    16.00       6 .  0000000355555677
     9.00       7 .  000345569
     2.00       8 .  00

  Stem width:     10.00
  Each leaf:       1 case(s)
```

Figure 15. Distribution of cervical right rotation measurements

```
Frequency     Stem &  Leaf

     4.00 Extremes    (=<25)
     5.00       3 .  00002
     6.00       3 .  555555
    18.00       4 .  000000000000000001
    21.00       4 .  555555555555555555558
    17.00       5 .  00000000000000334
     8.00       5 .  55555558
     1.00       6 .  0

  Stem width:     10.00
  Each leaf:       1 case(s)
```

Figure 16. Distribution of cerical left side-bending measurements

```
Frequency     Stem &  Leaf

      .00       1 .
     2.00       1 .  55
     1.00       2 .  0
     3.00       2 .  555
     8.00       3 .  00000003
    12.00       3 .  555555555566
    12.00       4 .  000000000344
    17.00       4 .  55555555555555558
    16.00       5 .  0000000000000012
     9.00       5 .  555555556

  Stem width:     10.00
  Each leaf:       1 case(s)
```

Figure 17. Distribution of cervical right side-bending measurements

```
Frequency     Stem &  Leaf

     1.00 Extremes    (=<30)
     2.00        3 .  44
    23.00        3 .  5566667777888999999999
    33.00        4 .  000000000111111111222223333333444
    19.00        4 .  5555556666667778999
     1.00        5 .  3
     1.00 Extremes    (>=55)

 Stem width:     10.00
 Each leaf:       1 case(s)
```

Figure 18. Distribution of wrist anterior-posterior measurements

```
Frequency     Stem &  Leaf

     1.00 Extremes    (=<40)
     1.00        4 .  7
    12.00        5 .  022222233444
    20.00        5 .  55555555666677777889
    33.00        6 .  000000000000011111122223333334444
    11.00        6 .  56777777888
     2.00        7 .  00

 Stem width:     10.00
 Each leaf:       1 case(s)
```

Figure 19. Distribution of wrist medial-lateral measurements

The hypothesized level of reliability for each clinical examination variable and the status of the hypothesis is listed in Figures 20 through 22. All judgments of hypotheses for provocative tests were based on tests with dichotomous results. In summary, the following levels of reliability according to the criteria of Landis and Koch[260] or Portney and Watkins[243p514] were determined for the clinical examination variables in this study:

Excellent (K>.75 or ICC>.90): Distraction, ULTT A and B (dichotomized), CCT (dichotomized), Phalen's test (dichotomized), questions 2-"Where most bothersome..", 4-"Hand fat/swollen..", 5-"Fumbling/dropping..", 8-"Night symptoms wake..", and 10-"Hand shaking improves.."

Fair to Good (K=.40 -- .75) or Good (ICC= .75 -- .90): Spurling's A and B, Valsalva, Tinel's A, ULTTA and B (collapsed), CCT, Phalen's (collapsed), questions 1-

"Most bothersome symptoms..", 3-"Symptom behavior..", 6-"Entire limb numb..", 7-"Symptoms keep from sleep..", 9-"Neck movement improves..", 11-"Worse with hand use..", C5, C7 and T1 dermatomes, MMT of the deltoid, biceps brachii, and extensor carpi radialis longus/brevis, bicep MSR, median sensory fields 1, 2, and 3, dermatome (any abnormality), myotome (any abnormality), and median sensory levels (any abnormality), cervical flexion, extension, left rotation, and the anterior-posterior/medial-lateral measurements of the wrist ratio.

Poor (K<.40) or Poor to Moderate (ICC<.75): Shoulder abduction, Tinel B, C6 and C8 dermatomes, MMT of the triceps brachii, flexor carpi radialis, abductor pollicis brevis, and the first dorsal interossei, cervical right rotation and bilateral sidebending.

CR Variables	Hypothesis status	CTS Variables	Hypothesis status
Questions 1-3, 6, 7, and 9	*Accepted*: question 2 *Rejected*: questions 1, 3, 6, 7, and 9	Questions 1-6, 8, 10, and 11	*Accepted*: questions 2, 4, 5, 8, and10 *Rejected*: questions 1, 3, and 11
Neck ROM	*Rejected*: All ROM parameters	Wrist Ratio measurements	*Rejected*: both anterior-posterior and medial-lateral measures
		Tinel's A	*Rejected*
		Tine's B	*Rejected*

Figure 20. Hypothesis: Clinical examination variables will demonstrate an Excellent level of reliability (Kappa > .75 or Intraclass Correlation Coefficient >.90)

CR Variables	Hypothesis status	CTS Variables	Hypothesis status
Sensation	*Accepted*: C5, C7, and T1 dermatome *Rejected*: C6 and C8 dermatomes	Sensation	*Accepted*: Median nerve sensory fields 1 - 3
MSR*	*Accepted*: biceps brachii	MMT abductor pollicus brevis**	*Accepted*
MMT	*Accepted*: deltoid, biceps brachii, extensor carpi radialis, *Rejected*: triceps brachii, flexor carpi radialis, adbuctor pollicus brevis*	Phalen's	*Rejected*
Spurling's A	*Accepted*	CCT	*Rejected*
Spurling's B	*Accepted*		
Shoulder Abduction	*Rejected*		
Valsalva	*Accepted*		
Neck Distraction	*Rejected*		

Figure 21. Hypothesis: Clinical examination variables will demonstrate a Fair to Good (K= .40 - .75) or Good (ICC= .75 - .90) level of reliability (*There was no variation across raters for triceps and brachioradialis reflex). **accepted when assessed with CTS subjects only

CR Variables	Hypothesis status	CTS Variables	Hypothesis status
ULTT A	*Rejected*	ULTT A	*Rejected*
ULTT B	*Rejected*	ULTT B	*Rejected*

Figure 22. Hypothesis: Clinical examination variables will demonstrate a Poor (K < .40) or Poor to Moderate (ICC <.75)level of reliability

Based on the 95CI's for each variable, the level of reliability was considered to be definitive for the following variables:

Excellent: question 5

Fair to Good (Kappa) or Good (ICC): Distraction, ULTTA (3 levels), ULTT A & B (dichotomous). CCT (3 levels), CCT and Phalen's (dichotomous), questions 1, 2, 4, 7, and 8 – 11, and wrist medial/lateral measurements.

No clinical examination variable had a poor level of reliability that was considered definitive.

6.3 Hypothesis #2 – Diagnostic Accuracy

6.3.1 Diagnostic Characteristics of Single Examination Items for CR and CTS

The first rater's results were used in all cases to determine the diagnostic accuracy of clinical examination variables. To calculate diagnostic accuracy characteristics for clinical examination variables, the following procedure was employed: For CR, the 13 subjects classified as CR formed the disease positive group. All subjects classified as normal or as CTS served as the control group. Likewise, for CTS, the 31 subjects classified as CTS formed the disease positive group and all subjects classified as normal or CR served as the control group. The 6 subjects with classifications other than normal, CR, or CTS were excluded from the analysis and the remaining 75 subjects were used for diagnostic accuracy calculations. For variables that had no false negative or false positive findings, .5 was added to each cell for adjustment.[271] The prevalence of CR and CTS for the total sample of 81 subjects was 16% and 38%, respectively. The diagnostic characteristics for predictor variables are shown by diagnostic category in Table 17 and 18. Values that met criteria for acceptability are in bold. Questions 1 – 3 are multi-level response items and do not have negative responses. Therefore, diagnostic characteristic are assigned to each level. The Likelihood ratio index (LRi) associated with each level is interpreted as a positive Likelihood ratios (LR+) because an absent or "negative" response for one level is the positive response of a different level.[54] Descriptive statistics for intial NDI scores and for the initial FSS and SSS scores of the CR and CTS groups, respectively, are listed in Table 19.

Table 17 Cervical Radiculopathy: Sensitivity (Sn), Specificity (Sp), Negative Likelihood Ratios (LR-), and Positive Likelihood Ratios (LR+) with 95% confidence intervals (95CI)

Variable	Sn 95CI	Sp 95CI	LR- 95CI	LR+ 95CI
question 1-"Most bothersome Sx's.				
i. Pain	.46 (.19 - .73)	.53 (.41 - .66)	*	.99 (.52 – 1.9)
ii. Numb/tingling	.46 (.19 - .73)	.55 (.42 - .67)		1.0 (.53 – 2.0)
iii. Loss of feeling	.08 (.00 - .92)	**.92** (.85 - .99)		.95 (.12 - 7.5)
question 2-"Where most bothersome.."				
i. Neck	.15 (.00 - .35)	**.92** (.85 - .99)		1.9 (.41 – 8.6)
ii. Shoulder/scap.	.46 (.19 - .73)	**.84** (.74 - .93)	*	**2.8** (1.2 – 6.5)
iii. Arm AE	**	**.93** (.87 – 1.0)		**
iv. Arm BE	**	**.84** (.74 - .93)		**
v. Hand or fingers	.38 (.12 - .48)	.48 (.35 - .60)		.73 (.35 – 1.5)
question 3-"Sx. behavior.."				
i. Constant	.08 (.00 - .22)	**.84** (.75 - .93)	*	.48 (.07 – 3.4)
ii. Intermittent	.56 (.06 - .56)	.63 (.51 - .75)		.83 (.34 – 2.0)
iii. Variable	**.88** (.35 - .88)	.53 (.41 - .66)		1.3 (.79 – 2.2)
question 6-"Entire limb numb.."	.23 (.00 - .46)	**.73** (.61 - .84)	1.1 (.76 – 1.5)	.84 (.29 – 2.5)
question 7-"Sx's. keep from sleep.."	.62 (.35 - .88)	**.90** (.82 - .99)	**.43** (.21 - .85)	**6.5** (2.3 - 18.0)
question 9-"Neck move improves.."	.69 (.44 - .94)	**.70** (.59 - .82)	**.44** (.19 – 1.0)	**2.4** (1.4 – 4.0)
C5 Dermatome	.15 (.00 - .35)	**.89** (82 - .97)	.95 (.74 – 1.2)	1.4 (.32 – 5.8)
C6 Dermatome	.15 (.00 - .35)	**.70** (.59 - .82)	1.2 (.90 – 1.6)	.52 (.14 – 2.0)
C7 Dermatome	**	**.74** (.63 - .85)	**	**
C8 Dermatome	**	**.82** (.72 - .92)	**	**
T1 Dermatome	**	**.82** (.72 - .91)	**	**
MMT deltoid	.23 (.00 - .46)	**.89** (.81 - .97)	.87 (.64 – 1.2)	**2.0** (.61 - 6.9)
MMT biceps brachii	.15 (.00 - .35)	**.94** (.87 – 1.0)	.90 (.71 – 1.2	**2.4** (.49 – 11.7)
MMT extensor carpi radialis longus/brevis	.08 (.00 - .22)	**.90** (.83 - .98)	1.02 (.86 – 1.2)	.79 (.10 – 6.1)

Table 17 (cont'd)

MMT triceps brachii	.08 (.00 - .22)	**.94** (.87 – 1.0)	.99 (.83 – 1.2)	1.2 (.14 – 9.8)
MMT flexor carpi radialis	**	**.90** (.83 - .98)	**	**
MMT abductor pollicus brevis	**	**.84** (.75 - .93)	**	**
MMT first dorsal interosseus		**.94** (.87 – 1.0)		
biceps brachii MSR	.23 (.00 - .46)	**.95** (.90 - 1.0)	.81 (.60 – 1.2)	**4.8** (1.1 – 21.0)
Spurling's A	.46 (.19 - .73)	**.87** (.78 - .95)	.62 (.37 – 1.03)	**3.5** (1.5 – 8.4)
Spurling's B	.46 (.19 - .73)	**.75** (.65 - .86)	.71 (.42 – 1.20)	1.9 (.90 – 3.9)
Shoulder Abduction	.08 (-.07 - .22)	**.92** (.85 - .99)	1.0 (.84 – 1.19)	.95 (.12 – 7.5)
Valsalva	.31 (.06 - .56)	**.94** (.87 – 1.0)	.74 (.51 – 1.1)	**4.8** (1.4 – 16.7)
Distraction	.38 (.12 - .65)	**.92** (.84 - .99)	.67 (.43 – 1.0)	**4.5** (1.5 – 13.4)
Upper limb tension test A[+]	**.96** (.87 – 1.0)	.23 (.12 - .34)	**.15** (.01 – 2.4)	1.3 (1.1 – 1.5)
Upper limb tension test B	.62 (.35 - .88)	.34 (.22 - .46)	1.1 (.52 – 2.5)	.93 (.58 – 1.5)
cervical flexion (<37)	.08 (-.07 - .22)	**.94** (.87 – 1.0)	.99 (.83 – 1.2)	1.2 (.14 – 9.8)
cervical extension (<55)	.38 (.13 - .65)	**.71** (.60 - .82)	.87 (.55 - .1.4)	1.3 (.60 – 2.9)
cervical left rotation (<48)	.11 (.00- .32)	**.81** (.64 - .98)	1.1 (.80 – 1.5)	.58 (.08 – 4.52)
Cervical left rotation (<57)	**.89** (.68 – 1.0)	.57 (.36 - .78)	**.19** (.03 – 1.4)	**2.07** (1.2 – 3.6)
visual analog scale worse (>7.5)	.45 (.16 - .75)	**.87** (.78 - .96)	.63 (.36 – 1.08)	**3.6** (1.38 – 9.2)

Table 18 Carpal Tunnel Syndrome: Sensitivity (Sn), Specificity (Sp), Negative Likelihood Ratios (LR-), and Positive Likelihood Ratios (LR+) with 95% confidence intervals (95CI)

Variable	Sn 95CI	Sp 95CI	LR- 95CI	LR+ 95CI
question 1-"Most bothersome Sx's.				
i. Pain	.35 (.19 - .52)	.45 (.31 - .60)	*	.65 (.38 – 1.1)
ii. Numb/tingling	.58 (.41 - .75)	.64 (.49 - .78)		1.6 (.98 – 2.6)
iii. Loss of feeling	.06 (-.02 - .15)	**.91** (.82 - .99)		.71 (.14 – 3.6)
question 2-"Where most bothersome.."				
i. Neck	.06 (-.02 - .15)	**.88** (.79 - .98)		.55 (.12 – 2.7)
ii. Shoulder/scap.	.16 (.03 - .29)	**.74** (.61 - .87)	*	.63 (.24 – 1.6)
iii. Arm AE	.06 (.00 - .15)	**.95** (.89 – 1.0)		1.39 (.21 – 9.3)
iv. Arm BE	.10 (.00 -.20)	**.84** (.73 - .95)		.59 (.17 – 2.1)
v. Hand or fingers	.61 (.44 - .78)	.58 (.43 - .73)		1.5 (.93 – 2.3)
question 3-"Sx. behavior.."				
i. Constant	.23 (.08 - .37)	**.91** (.82 - .99)	*	**2.5** (.79 – 7.8)
ii. Intermittent	.42 (.25 - .59)	.68 (.54 - .82)		1.3 (.72 – 2.4)
iii. Variable	.35 (.19 - .52)	.41 (.26 - .55)		.60 (.35 – 1.0)
question 4-"Hand fat/swollen.."	.48 (.31 - .66)	.66 (.52 - .80)	.78 (.52 – 1.2)	1.42 (.82 – 2.5)
question 5-"Fumble and dropping.."	**.74** (.59 - .90)	.61 (.47 - .76)	**.42** (.22 - .80)	1.9 (1.3 – 1.9)
question 6-"Entire limb numb.."	.35 (.19 - .52)	**.80** (.68 - .91)	.81 (.60 – 1.1)	1.7 (.82 – 3.7)
question 8-"Night sx's. wake.."	.56 (.41 - .71)	.32 (.18 - .46)	1.4 (.79 – 2.4)	.82 (.59 – 1.2)
question 10-"Hand shaking improves."	**.77** (.63 - .92)	.60 (.46 - .75)	**.37** (.19 - .75)	**2.0** (1.3 – 3.0)
question 11-"Worse with hand use.."	**.74** (.59 - .90)	.39 (.24 - .53)	.67 (.33 – 1.4)	1.2 (.88 – 1.7)
Median sensory field 1	.55 (.37 - .72)	.68 (.54 - .82)	.66 (.53 – 1.0)	1.7 (1.0 – 3.0)
Median sensory field 2	.57 (.29 - .65)	.65 (.59 - .78)	.84 (.56 – 1.3)	1.3 (.74 – 2.2)

Table 18 (cont'd.)

Median sensory field 3	.43 (.26 - .61)	**.75** (.62 - .88)	.76 (.53 – 1.2)	1.7 (.90 – 3.3)
MMT abductor pollicus brevis	.19 (.95 - .33)	**.91** (.82 - .99)	.89 (.73 – 1.1)	**2.1** (.66 – 6.9)
Tinel's A	.42 (.25 - .59)	.56 (.41 - .71)	1.0 (.70 – 1.6)	.95 (.56 – 1.6)
Tinel's B	.48 (.66 - .31)	.66 (.52 - .80)	.78 (.52 – 1.2)	1.4 (.82 – 2.5)
upper limb tension test A	**.71** (.55 - .88)	.12 (.02 - .22)	2.4 (.87 – 6.59)	.81 (.63 – 1.1)
upper limb tension test B	.59 (.41 - .77)	.20 (.17 - .44)	1.4 (.73 – 2.6)	.84 (.58 – 1.2)
carpal compression test	**.74** (.59 - .90)	.47 (.32 - .61)	.55 (.28 – 1.1)	1.4 (.98 – 2.0)
Phalen's test	**.70** (.54- .86)	.37 (.23 - .52)	.81 (.41 -1.6)	1.1 (.80 – 1.6)
wrist ratio (>.68)	**.71** (.55 - .87)	.63 (.47 - .78)	.68 (.41 – 1.1)	1.5 (.92 – 2.6)
wrist ratio (>.70)	.58 (.39 - .77)	.63 (.47 - .78)	.68 (.41 – 1.1)	1.5 (.92 – 2.6)
Hand diagram	.68 (.51 - .84)	.42 (.27 - .57)	.77 (.41 – 1.4)	1.2 (.82 – 1.7)
Symptom Severity Scale (>3.2)	.35 (.19 - .52)	**.86** (.76 - .96)	.75 (.56 – 1.0)	**2.5** (1.1 – 6.1)
Functional Status Scale (>2.5)	.42 (25 - .59)	**.86** (.76 - .97)	.67 (.49 - .93)	**3.1** (1.3 – 7.2)

Table 19 Descriptive Statistics for initial patient self-report measures:

Measure	Number of Subjects		Mean		Minimum		Maximum		Std. Deviation	
	CR	CTS	CR	CTS	CR	CTS	CR	CTS	CR	CTS
NDI	13	28	25%	21%	2%	.00	64%	20%	19.9	14.6
SSS	13	31	2.3	2.8	1.0	1.0	3.6	4.1	.9	.8
FSS	13	31	1.7	2.2	1.0	1.0	3.9	3.5	1.0	.8
Valid N	13	28								

NDI- Neck Disability Index, SSS- Symptom Severity Scale, FSS- Functional Status Scale

All diagnostic accuracy characteristics of provocative tests were computed using dichotomous ratings only. Receiver operating characteristic curves were generated for continuous variables to establish optimum cut-off points and are listed in Figures 23 through 28. Receiver operating curves that revealed no potentially useful cut-off points for the following CR variables are not shown: NDI, cervical right rotation, bilateral sidebending, and the VASnow pain rating. The area under the curves for each of these variables was less than .54. Receiver operating curves revealed no potentially useful cut-off values for the following CTS variables and are not shown: VASnow and VASworse pain ratings. The area under the curves for each of these variables was less than .42. Because the proposed best cut-off point for the wrist ratio is.70,[174] an ROC curve analysis was used to determine whether this was the best cut-off value in this sample of patients (Figure 29. The best wrist ratio cut-off point for this sample of patients was .68 with the area under the curve= .58.

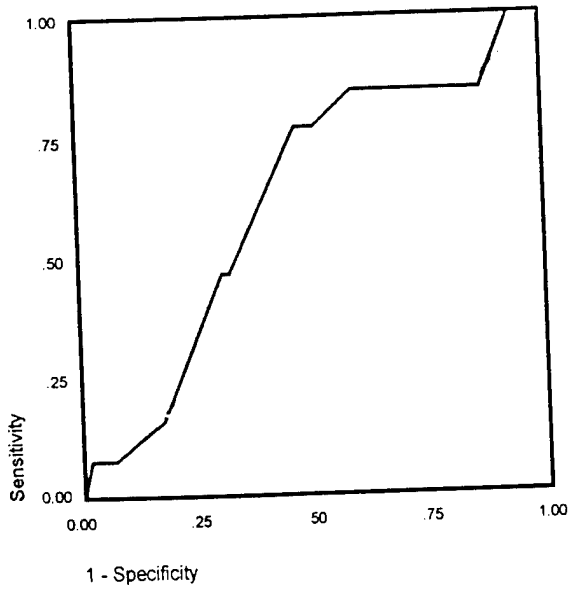


Figure 23. Receiver operating curve for cervical flexion: Cut-off 37 degrees

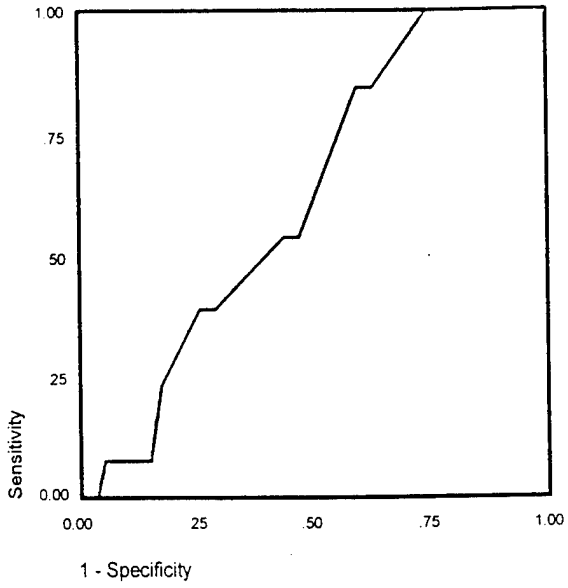


Figure 24. Receiver operating curve for cervical Extension: Cut-off 55 degrees

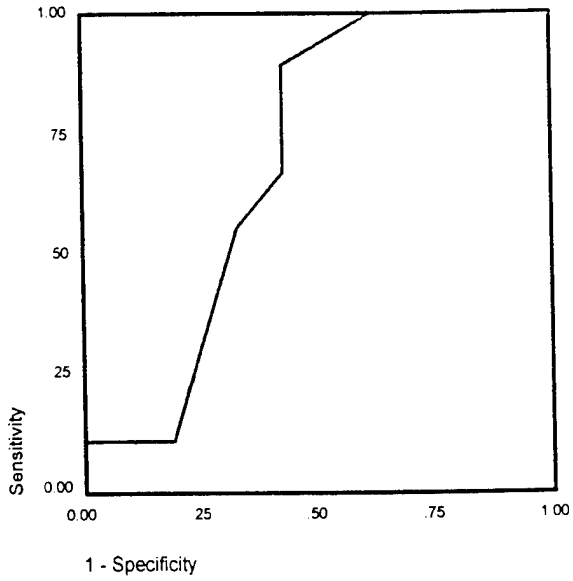


Figure 25. Receiver operating curve for cervical left rotation: Cut-offs <48 and <57 degrees

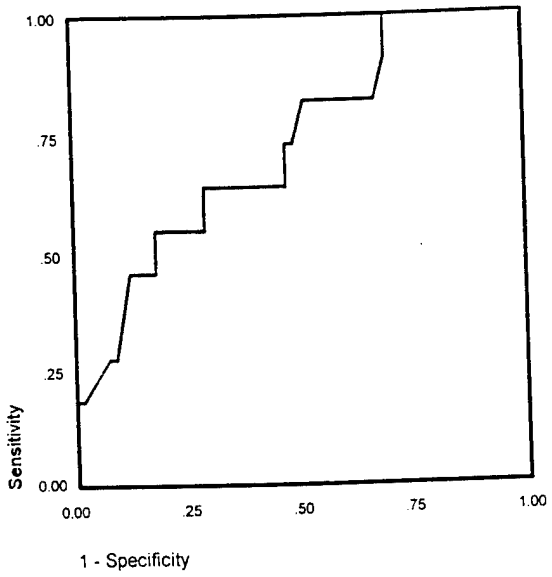


Figure 26. Receiver operating curve for visual analog "worst" pain rating (cervical radiculopathy): Cut-off >7.5

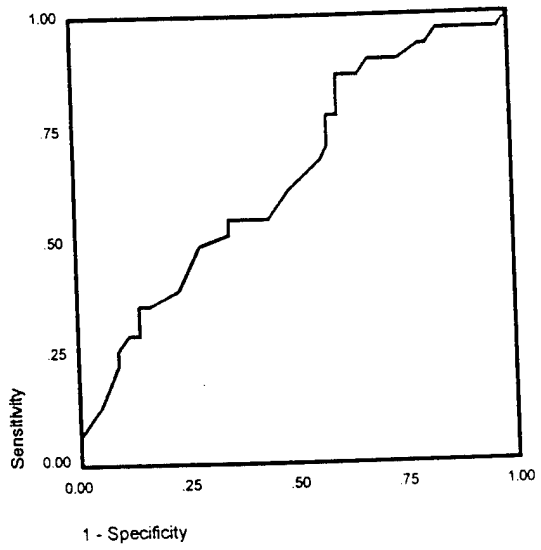


Figure 27. Receiver operating curve for Symptom Severity Scale: cut-off >3.2.

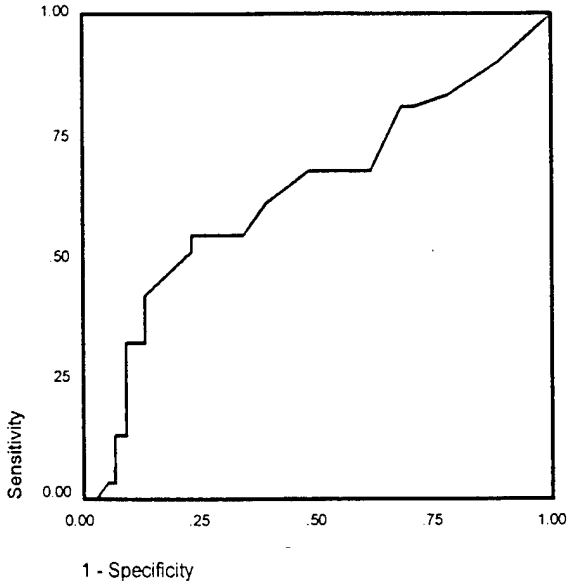


Figure 28. Receiver operating curve for Functional Status Scale: cut-off >2.5.

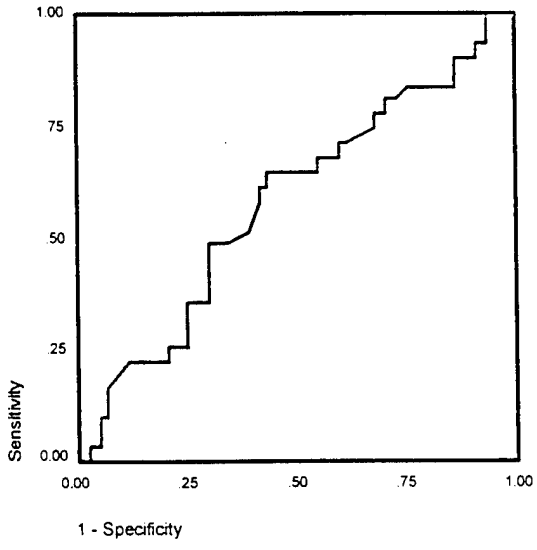


Figure 29. Receiver operating curve for wrist ratio: cut-off >.68.

The hypothesized levels of diagnostic accuracy for each clinical examination variable and the status of the hypotheses are listed in Figures 30 - 33.

CR Variables	Hypothesis status
question 2-"Where most bothersome.."	*Accepted*: all levels except v.
question 7-"Symptoms keep from sleep.."	*Accepted*
Sensation	*Accepted*: all dermatome levels
Muscle stretch reflexes	*Accepted*: biceps brachii, Undeterminable: brachiradialis, triceps brachii
MMT	*Accepted*: all muscles
Distraction	*Accepted*
Spuling's A	*Accepted*
Spurling's B	*Accepted*
Valsalva	*Accepted*
shoulder abduction	*Accepted*

Figure 30 Hypothesis: Cervical radiculopathy clinical examination variables will demonstrate an acceptable level of diagnostic accuracy (Sn or Sp ≥ .70 or LR+ ≥ 2.0 or LR- ≤ .50)

CR Variables	Hypothesis status
question 1-"Most bothersome	*Accepted:* levels i and ii
symptoms.."	*Rejected:* level iii
question 3-"Symptom behavior.."	*Accepted*: level ii
	Rejected: levels i and iii
question 6-"Entire limb numb.."	*Rejected*
question 9-"Neck movement improves"	*Rejected*
upper limb tension test A	*Rejected*
upper limb tension test B	*Accepted*
cervical range-of-motion	*Accepted*: right rotation and bilateral sidebending *Rejected*: flexion, extension, left rotation
Visual analog scale	Rejected
Neck Diability Index	*Accepted*

Figure 31. Hypothesis: Cervical radiculopathy clinical examination variables will not demonstrate an acceptable level of diagnostic accuracy

Variables	Hypothesis status
question 2-"Where most bothersome.."	*Accepted:* levels i, ii, iii, and iv *Rejected*: *level v*
question 4-"Hand fat/swollen..:	*Rejected*
question 5-"Fumbling/dropping.."	*Accepted*
question 8-"Night symptoms wake.."	*Rejected*
question 10-"Hand shaking improves.."	*Accepted*
question 11-"Worse with hand use.."	*Accepted*
Phalen's test	*Accepted*
carpal compression test	*Accepted*
Tinel's A	*Rejected*
Tinel's B	*Rejected*
wrist ratio	*Rejected*
hand diagram	*Rejected*

Figure 32. Hypothesis: Carpal tunnel syndrome clinical examination variables will demonstrate an acceptable level of diagnostic accuracy (Sn or Sp ≥ .70 or LR+ ≥ 2.0 or LR- ≤ .50).

Variables	Hypothesis status
question 1-"Most bothersome symptoms.."	*Accepted:* levels i, ii *Rejected*: level iii
question 3-"Symptom behavior.."	*Accepted:* levels ii and iii *Rejected*: level i
question 6-"Entire limb numb.."	*Rejected*
Manual muscle test (abductor pollicus brevis)	*Rejected*
Sensation	*Accepted*: median nerve field1, median nerve field2 *Rejected:*, median nerve field3
Visual analog scale	*Accepted*
Symptom Severity Scale	*Rejected*
Functional Status Scale	*Rejected*

Figure 33 Hypothesis: Carpal tunnel syndrome clinical examination variables will not demonstrate an acceptable level of concurrent validity

Based on the lower bound of the 95CI, definitive findings for the Sp and Sn are listed in Figure 34. Variables that did not have any true positive or true negative results were excluded. Although definitive results were found to occur, only 18 variables had adequate power and all pertained to Sp values only. Sixty-two subjects classified as positive for the disease of interest are required to achieve a 95CI for a Sn of .80 with a lower limit of .70. The same number of non-diseased subjects is required to achieve a 95CI for a Sp of .80 with a lower limit of .70. With the exception of question 7 (LR+=6.5, 95CI= 2.31 - 18.0), decisions regarding definitively unacceptable LR's were not made due a lack of power for achieving the previously specified upper and lower 95CI limits.

Cervical radiculopathy Variables		Carpal tunnel syndrome Variables	
Sensitivity	Specificity	Sensitivity	Specificity
question 7	question 1 (iii) question 2(i – iv) questions 7		question 1(iii) question 2(i & iii) question 3(i)
upper limb tension test A	Spurling's A		symptom Severity Scale
question 3(iii)	shoulder Abduction Valsalva		Functional Status Scale abductor pollicus brevis muscle test
	distraction upper limb tension test A cervical flexion		
	visual analog scale (worse) C5 dermatome level biceps brachii muscle stretch reflex all muscle tests		

Figure 34. Definitively acceptable sensitivity and specificity findings of clinical examination variables

Cervical radiculopathy Variables		Carpal tunnel syndrome Variables	
Sensitivity	Specificity	Sensitivity	Specificity
question 2(i and v)	Question 1(i and ii)	question 1(i and iii)	question 3(iii)
question 3(i and ii)	Question 2(v)	question 2(i, ii, iii, and iv)	question 8
question 6	Question 3(iii)	question 3(i – iii)	question 11
		questions 4	
		question 6	
C5 and C6	upper limb tension	Median sensory field 2	Phalen's test
dermatomes	test A and B	Median sensory field 3	carpal compression test upper limb tension test A and B
biceps brachii		abductor pollicus brevis	hand diagram
muscle stretch reflex		muscle test	
All muscle tests		Tinel's A and B	wrist ratio
Valsalva		Functional Status Scale	
shoulder abduction		Symptom Severity Scale	
distraction			
cervical flexion, extension, left rotation			

Figure 35. Definitively unacceptable sensitivity and specificity findings of clinical examination variables

Figure 36 summarizes clinical examination variables with acceptable Likelihood ratios for each by each respective condition.

Acceptable Likelihood ratios (LR+ ≥ 2.0 or LR- ≤ .50)					
CR Variables			**CTS Variables**		
	LR-	LR+/LRi		LR-	LR+/LRi
Questions:			Questions:		
2(ii)		2.8	3(i)		2.5
7	.43	6.5	5	.42	
9	.37	2.4	10	.37	2.0
Neurological Examination			Neurological Examination		
deltoid muscle			abductor pollicus		
muscle test		2.0	brevis muscle test		2.1
biceps brachii					
Muscle test		2.4			
biceps muscle					
Stretch reflex		4.8			
Self-report			Self-report		
visual analog			SSS (>3.2)		2.5
scale(worse)		3.6			
			FSS (>2.5)		3.1
Provocative Tests					
ULTT A	.15				
Valsalva		4.8			
distraction		4.5			
Spurling's A		3.5			
Scaled measurements					
Involved cervical					
Rotation (left only)	.19	2.1			

Figure 36. Summary of acceptable Likelihood ratios

6.3.2 Diagnostic Characteristics of Single Examination Items for CTS Subclassified Groups

To assess the impact of spectrum bias on test sensitivity, two groups were formed from the CTS diagnosed subjects. Group A consisted of subjects subclassified as mild/moderate and group B consisted of subjects subclassified as pronounced/severe. Subjects classified as normal or CR comprised the control group; groups A and B were excluded from each other's respective control group. Descriptive statistics for the initial SSS, FSS and VASnow scores of the two groups are listed in Table 20. Because the control group remained unchanged, only Sn values were affected. Therefore, only Sn and Likelihood ratio values for the two groups are reported in Tables 21 – 22. Calculations were performed as previously described and for variables that had no false negative or false positive findings, .5 was added to each cell for adjustment.[271] The cut-off values previously determined for all CTS group were used and no ROC curves or separate cut-off values were established for the two groups.

Numerical differences in diagnostic characteristics between the groups that may indicate a trend are in bold type if the value was considered acceptable and exceeded the following difference thresholds: Sn= > .10, LR-= > .15, and LR+= >1.5. Power was not satisfactory to detect a difference for the previously specified levels of acceptance. Question 5(v) for the pronounce/severe group was the only variable with a significantly different Sn value as determined by the 95CI interval and is italicized in both tables.

Table 20. Descriptive statistics of self-report instruments for subjects by group severity

Measure	Group	N	Mean	Std. Deviation
SSS1	Mild/moderate	21	2.82	.86
	Pronounced/Severe	10	2.80	.42
FSS1	Mild/moderate	21	2.35	.85
	Pronounced/Severe	10	1.99	.72
VASNOW1	Mild/moderate	19	2.8	2.5
	Pronounced/Severe	8	2.4	2.4

Table 21. Mild/Moderate Carpal Tunnel Syndrome: Sensitivity (Sn), Negative Likelihood Ratio (LR-), and Positive Likelihood Ratio (LR+) of clinical examination variables with 95% confidence intervals (95CI).

Variable	Sn 95CI	LR- 95CI	LR+ 95CI
question 1-"Most bothersome Sx's.			
i. Pain	.43 (22 - .64)	*	.79 (.45 – 1.4)
ii. Numb/tingling	**.52 (.31 - .74)**		1.4 (.82 – 2.5)
iii. Loss of feeling	.05 (.00 - .14)		.52 (.06 – 4.4)
question 2-"Where most bothersome.."			
i. Neck	10 (.00 - .22)		.82 (.17 – 3.9
ii. Shoulder/scap.	.19 (.02 - .36)	*	.74 (.27 – 2.1)
iii. Arm AE	.10 (.00 - .22)		**2.1 (.31 – 13.5)**
iv. Arm BE	.14 (.00 - .29)		.88 (.25 – 3.1)
v. Hand or fingers	***.48 (.26 - .69)***		1.1 (.64 – 2.0)
question 3-"Sx. behavior.."			
i. Constant	.14 (.00 - .29)		**1.6 (.39 – 6.4)**
ii. Intermittent	.48 (.26 - .69)	*	1.6 (.80 – 2.8)
iii. Variable	.38 (.17 - .59)		.64 (.35 – 1.2)
question 4-"Hand fat/swollen.."	.48 (.26 - .69)	.79 (.50 – 1.3)	1.4 (.76 – 2.6)
question 5-"Fumble and dropping.."	**.67 (.47 - .87)**	**.54 (.28 – 1.0)**	1.7 (1.1 – 2.7)
question 6-"Entire limb numb.."	.38 (.17 - .59)	.78 (.54 – 1.1)	1.9 (.84 – 4.1)
question 8-"Night sx's. wake.."	.76 (.58 - .76)	.75 (.31 – 1.8)	1.1 (.82 – 1.5)
question 10-"Hand shaking improves."	.76 (.58 - .76)	.39 (.18 - .88)	1.9 (1.2 – 3.0)
question 11-"Worse with hand use.."	.76 (.58 - .76)	.62 (.26 – 1.4)	1.2 (.89 – 1.7)
Median sensory field 1	**.43 (.22 - .64)**	**.84 (.55 – 1.3)**	1.4 (.70 – 2.6)
Median sensory field 2	.40 (.19 - .61)	.94 (.62 – 1.4)	1.1 (.57 – 2.1)
Median sensory field 3	.40 (.19 - .61)	.80 (.54 – 1.2)	1.6 (.76 – 3.4)
MMT abductor pollicus brevis	.19 (.02 - .36)	.89 (.71 – 1.1)	2.1 (.58 – 7.6)

Table 21 (cont'd.)

Tinel's A	.38 (.17 - .59)	1.1 (.72 – 1.7)	.86 (.45 – 1.6)
Tinel's B	.48 (.26 - .69)	.79 (.50 – 1.3)	1.4 (.76 – 2.6)
upper limb tension test A	.78 (.59 - .97)	1.9 (.57 – 6.2)	.88 (.67 – 1.2)
upper limb tension test B	.58 (.36 - .80)	1.4 (.69 – 2.8)	.83 (.54 – 1.3)
carpal compression test	.71 (.52 - .91)	.61 (2.9 – 1.3)	1.3 (.91 – 12.0)
Phalen's test	.70 (.50 - .90)	.81 (.37 – 1.8)	1.1 (.77 – 1.6)
wrist ratio (>.68)	**.76 (.68 - .94)**	.62 (.26 – 1.4)	1.2 (.89 – 1.7)
wrist ratio (>.70)	.62 (.41 - .83)	.67 (.37 – 1.2)	1.4 (.89 – 2.3)
Hand diagram	.67 (.47 - .87)	.80 (.40 – 1.6)	1.2 (.77 – 1.7)
Symptom Severity Scale (>3.2)	.43 (.22 - .64)	.66 (.45 - .98)	**3.1 (1.3 – 7.5)**
Functional Status Scale (>2.5)	.48 (.26 - .69)	.61 (.40 - .93)	3.5 (1.5 – 8.3)

Table 22 Prounounced/severe Carpal Tunnel Syndrome: Sensitivity (Sn), Negative Likelihood Ratio (LR-), and Positive Likelihood Ratio (LR+) of clinical examination variables with 95% confidence intervals (95CI)

Variable	Sn 95CI	LR- 95CI	LR+ 95CI
question 1-"Most bothersome Sx's.			
i. Pain	.20 (-.06 - .45)		.37 (.10 – 1.3)
ii. Numb/tingling	**.70 (.42 - .98)**	*	1.9 (1.1 – 3.4)
iii. Loss of feeling	.10 (-.09 - .29)		1.1 (.14 – 8.8)
question 2-"Where most bothersome.."			
i. Neck	**		**
ii. Shoulder/scap.	.40 (-.09 - .29)	*	.39 (.06 – 2.7)
iii. Arm AE	**		**
iv. Arm BE	**		**
v. Hand or fingers	***.90 (.71 – 1.0)***		2.15 (1.4 – 3.2)
question 3-"Sx. behavior.."			
i. Constant	.40 (.10 - .70)		**4.4 (1.3 – 14.7)**
ii. Intermittent	.30 (.02 - .58)	*	.94 (.33 –2.7)
iii. Variable	.30 (.02 - .58)		.51 (.19 – 1.4)

Table 22 (cont'd.)

question 4-"Hand fat/swollen.."	.50 (.19 - .81)	.76 (.39 –1.5)	1.5 (.70 – 3.1)
question 5-"Fumble and dropping.."	**.90 (.71 – 1.0)**	**.16 (.03 – 1.1)**	2.3 (1.5 - 3.6)
question 6-"Entire limb numb.."	.30 (.02 - .58)	.88 (.57 – 1.4)	1.5 (.48 – 4.5)
question 8-"Night sx's. wake.."	.70 (.42 - .98)	.94 (.33 – 2.7)	1.0 (.65 – 1.6)
question 10-"Hand shaking improves."	.80 (.55 – 1.1)	.33 (.09 – 1.2)	2.0 (1.3 – 3.3)
question 11-"Worse with hand use.."	.70 (.42 - .98)	.78 (.28 – 2.2)	1.1 (.71 – 1.8)
Median sensory field 1	**.80 (.55 – 1.0)**	**.29 (.08 – 1.0).**	2.5 (1.5 – 4.3)
Median sensory field 2	.60 (.30 - .90)	.63 (.28 – 1.4)	1.7 (.87 - 3.1)
Median sensory field 3	.50 (.19 - .81)	.67 (.35 – 1.3)	2.0 (.90 – 4.5)
MMT abductor pollicus brevis	.20 (-.05 - .45)	.88 (.64 - .88)	2.2 (.47 – 10.4)
Tinel's A	.50 (.19 - .81)	.90 (.46 – 1.8)	1.1 (.56 – 2.3)
Tinel's B	.50 (.19 - .81)	.76 (.39 – 1.5)	1.5 (.70 – 3.1)
upper limb tension test A	.60 (.30 - .0)	3.4 (1.1 – 10.3)	.68 (.41 – 1.1)
upper limb tension test B	.60 (.10 - .90)	1.3 (.55 – 3.2)	.86 (.50 – 1.5)
carpal compression test	.80 (.55 – 1.0)	.73 (.12 – 1.6)	1.5 (.99 – 2.3)
Phalen's test	.70 (.42 - .98)	.81 (.29 – 2.2)	1.1 (.70 – 1.8)
wrist ratio (>.68)	**.60 (.10 - .90)**	1.0 (.44 – 2.4)	.98 (.56 - .17)
wrist ratio (>.70)	.60 (.10 - .90)	.70 (.32 – 1.6)	1.4 (.76 – 2.6)
Hand diagram	.70 (.42 - .98)	.72 (.26 – 2.0)	1.9 (.75 – 1.9)
Symptom Severity Scale (>3.2)	.20 (-.05 - .45)	.93 (.67 – 1.3)	**1.4 (.34 – 6.1)**
Functional Status Scale (>2.5)	.30 (.02 - .58)	.81 (.53 – 1.2)	2.2 (.66 – 7.3)

6.4 Hypothesis #3 – Predictive Validity

6.4.1 Predictive Validity of Single Examination Items for CR Subjects:

The predictive validity of clinical examination variables for CR could not be meaningfully measured due to low prevalence. Ten out of 13 (77%) subjects returned

follow-up forms (mean follow-up days = 42.8, sd= 4.5). The surgical status was reported for nine of these subjects; none received surgery. Seven subjects responded when asked whether they had been offered surgery for their condition. Two of these subjects had been offered surgery. Global rating of change scores were available for ten out of 13 subjects. Based on a GRCS criteria of ≥ +/- 4points, four of these subjects improved and two were worse; one of the two subjects who had improved had been offered surgery. The descriptive statistics for NDI scores of subjects who improved are compared with the unimproved group and are listed in Table 23. No further analyses of these subjects were performed.

Table 23. Descriptive statistics of Neck Disablity Scores at follow-up.

Group	N	Mean: F/U and Chng	Standard Deviation: F/U and Chng	Minimum: F/U and Chng	Maximum: F/U and Chng
Improved	4	6.0/12.5	5.817.3	0/-4	14/36
Unimproved	6	31.8/-4.5	21.0/7.0	11/-14	66/4.0

F/U= Follow-up
Chnge= Change

6.4.2 Predictive Validity of Single Examination Items for CTS Subjects.

To measure the predictive validity of clinical examination variables for CTS subjects, diagnostic accuracy characteristics were calculated using the same method described for hypothesis #2, except selected measures of outcome were used as the gold standard. The results from the first rater were used in all cases to determine the predictive validity of clinical examination variables. Outcome variables included surgical status and patient perception of change based on a GRCS. A change score of ≥ 4 points on the GRCS was used as the cut-off used to classify a patient as improved and a change score of ≤ 4 was used to classify a patient as worsened. Patients who did not meet the directional change

criteria served as the control group. In addition to the clinical examination and self-report variables, the following measures were used included as predictor variables: FABQA, FABQB, median palmar sensory and motor latencies, median palmar sensory and motor amplitudes, EMG spontaneous activity rating of the abductor pollicus brevis muscle, and duration of symptoms.

Twenty-five of the 31 subjects (81%) classified with CTS returned follow-up forms and indicated whether they had received surgery (mean follow-up days = 59.0, sd= 6.9). Five subjects from two centers received surgery for their condition, one from the University of Pittsburgh Medical Center and four from WHMC. Three of these subjects had improved and one subject had worsened, and one was unchanged. An analysis of change for the 20 CTS subjects who had not received surgery using the criteria of $\geq$ +/- 4 points resulted in three subjects who improved and four who worsened. Due to the low prevalence of change in a single direction (improved or worsened), a revised criteria of $\geq$ +/- 3 points was used. Using the revised change criteria, five (25%) subjects improved and seven (35%) subjects worsened. Fourteen of the CTS subjects who had not received surgery responded when asked whether they had been offered surgery. Six of these 14 subjects had been offered surgery; two had improved and two had worsened. The descriptive statistics for SSS and FSS scores of subjects who changed are compared in Tables 24 and 25.

Table24 Descriptive statistics of self-report scores for improved and control subjects

Self-report Instrument	N	Group	Mean: F-U/Chng	Standard Deviation: F-U/Chng	Minimum: F-U/Chng	Maximum: F-U/Chng
Symptom	5	Improved	2.7/.58	.77/.66	1.6/-.28	3.6/1.5
Severity Scale	20	Unimproved	2.5/.04	.87/.47	1.0/-1.4	3.7/.64
Funtional	20	Improved	2.4/.53	.60/.44	1.75/.00	3.0/1.1
Status Scale	20	Unimproved	2.1/-.17	.80/.62	1.0/-1.5	3.3/.70

F-U= Follow-up
Chnge= Change

Table25 Descriptive statistics of self-report scores for worsened and control subjects

Self-report Instrument	N	Group	Mean: F-U/Chng	Standard Deviation: F-U/Chng	Minimum: F-U/Chng	Maximum: F-U/Chng
Symptom	7	Worsened	3.0/.18	.52/.36	2.3/-.37	3.6/.60
Severity Scale	13	Not Worsened	2.4/.17	.89/.66	1.0/-1.4	3.7/1.5
Funtional	7	Worsened	2.7/-.18	.53/.73	2.0/-1.5	3.3/.70
Status Scale	13	Not Worsened	1.8/.11	.70/.62	1.0/-.14	3.0/1.1

F-U= Follow-up
Chnge= Change

6.4.2.1 Surgical Intervention Gold Standard.

The five subjects who received surgery were considered positive for the condition and the remaining 20 subjects served as the control group. The prevalence of surgical intervention for the 25 subjects whose return forms were available was 20%.

All predictive validity characteristics of provocative tests were computed using dichotomous ratings only. ROC curves were generated for all continuous variables to identify the optimum cut-off points and are listed by diagnosis in Figures 37 – 43. Receiver operating curves revealed no potentially useful cut-off values for the following variables and are not shown: VASnow, VASworse, FABQ A, FABQ total, and median sensory amplitude. The area under the curve for each of these variables was less than .56. The predictive validity characteristics for the predictor variables are shown in Table 26. Values that met levels of acceptability are in bold. Items with no true positive or true negative responses are indicated by a double asterisk(**).

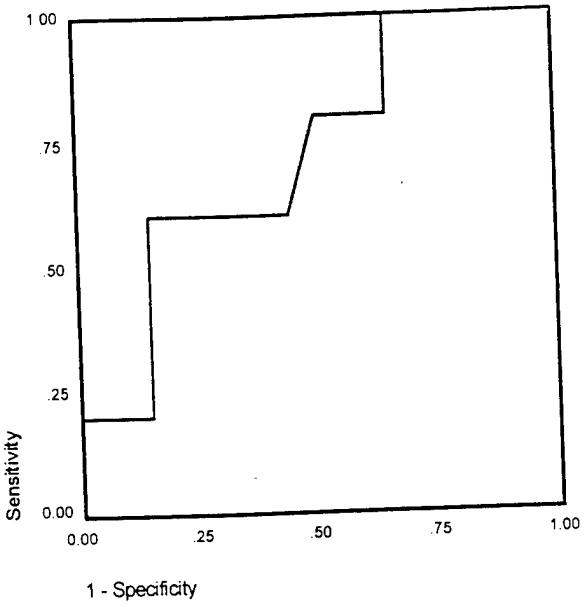


Figure 37. Receiver Operating Curve for wrist ratio: Cut-off >.73 degrees

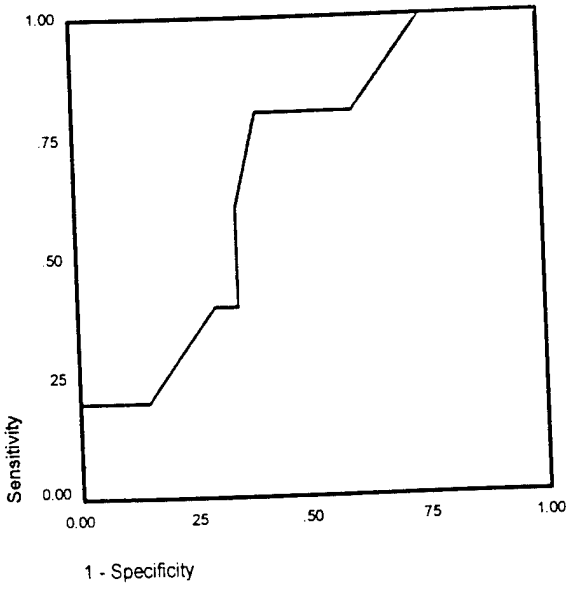


Figure 38. Receiver Operating Curve for Functional Status Scale: Cut-off >2.3 points

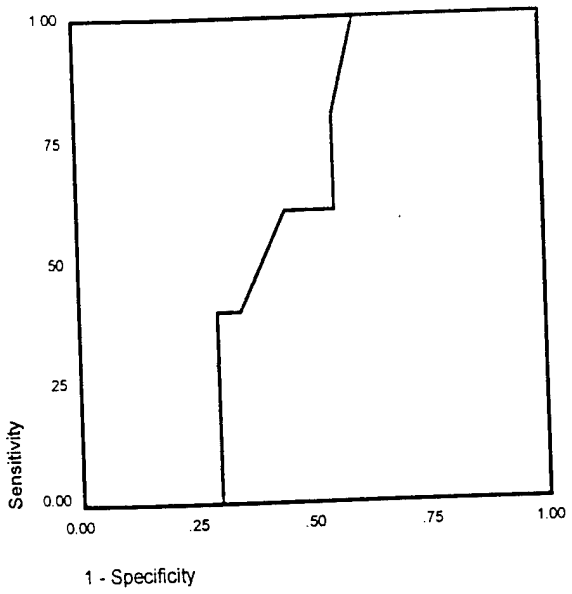


Figure 39. Receiver Operating Curve for Fear Avoidance Behavior Questionnaire B: Cut-off >31 points

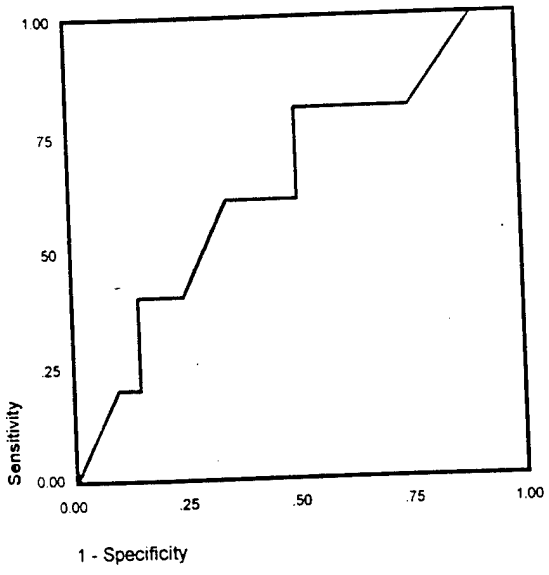


Figure 40. Receiver Operating Curve for involved median palmar sensory latency: Cut-off >.30 milliseconds

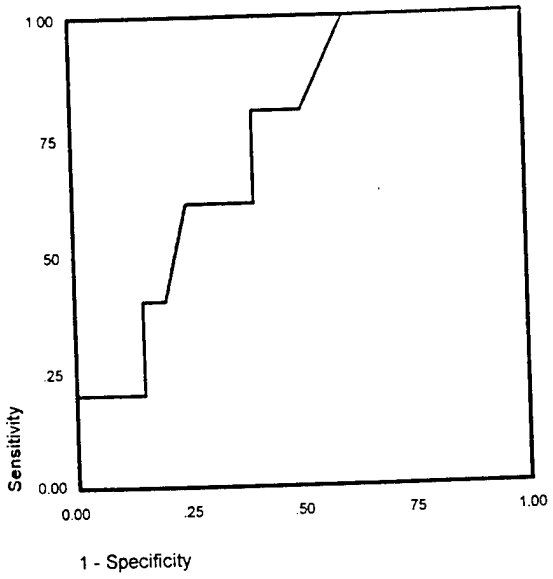


Figure 41. Receiver Operating Curve for involved median motor latency: Cut-off >5.0 milliseconds

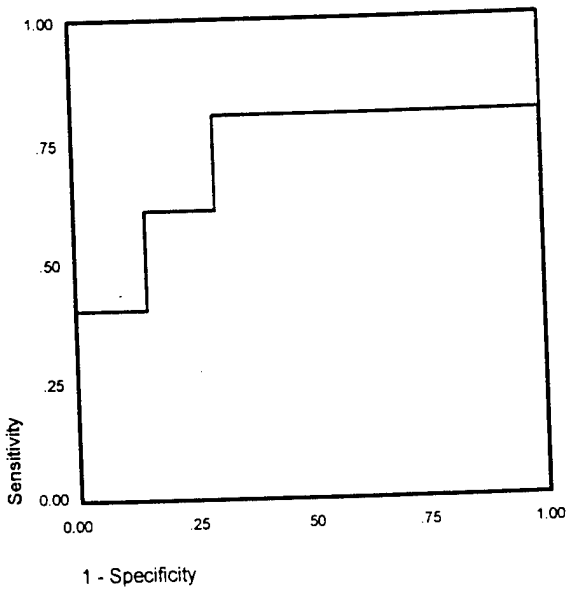


Figure 42. Receiver Operating Curve for involved median motor amplitude: Cut-off <4800 microvolts

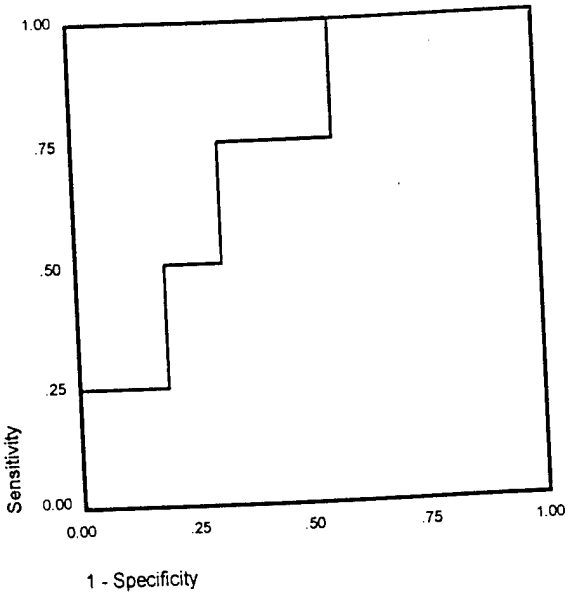


Figure 43. Receiver Operating Curve for duration of symptoms: Cut-offs <78 and >391 days

Table 26 Surgical intervention predictors for carpal tunnel syndrome subjects: Sensitivity (Sn), Specificity (Sp), Negative Likelihood Ratio (LR-), and Positive Likelihood Ratio (LR+) of clinical examination variables with 95% confidence intervals (95CI)

Variable	Sn	95CI	Sp	95CI	LR-	95CI	LR+	95CI
question 1-"Most bothersome Sx's.								
i. Pain	.40	(.00 - .83)	.55	(.33 - .77)			.89	(.27 – 2.9)
ii. Numb/tingling	.40	(.00 - .83)	.45	(.23 - .67)	*		.73	(.23 – 2.3)
iii. Loss of feeling	.25	(.00 - .60)	**.98**	(.91 – 1.0)			10.5	(.49 – 206.1)
question 2-"Where most bothersome.."								
i. Neck	**		**.90**	(.77 – 1.0)			**	
ii. Shoulder/scap.	**		**.75**	(.56 - .75)	*		**	
iii. Arm AE	**		**.95**	(.85 – 1.0)			**	
iv. Arm BE	**		**.90**	(.77 – 1.0)			**	
v. Hand or fingers	**.92**	(.70 – 1.0)	.50	(.29 - .50)			1.8	(1.1– 3.0)
question 3-"Sx. behavior.."								
i. Constant	.40	(.00 - .83)	**.79**	(.61 - .97)			1.9	(.48 – 7.6)
ii. Intermittent	.20	(.15 - .55)	.65	(.44 - .86)	*		.57	(.09 – 3.6)
iii. Variable	.40	(.00 - .83)	.55	(.33 - .77)			.89	(.27 – 2.9)
question 4-"Hand fat/swollen.."	40	(.00 - .83)	.50	(.28 - .50)	1.2	(.52 – 2.8)	.80	(.25 – 2.6)
question 5-"Fumble and dropping.."	**.92**	(.70 – 1.0)	.26	(.07 - .45)	**.32**	(.02 – 5.0)	1.3	(.99 – 1.7)
question 6-"Entire limb numb.."	.40	(.00 - .83)	.65	(.44 - .86)	.92	(.42 – 2.0)	1.1	(.33 – 3.9)
question 8-"Night sx's. wake.."	.60	(.17 – 1.0)	.30	(.10 - .50)	1.3	(.38 – 4.7)	.86	(.40 – 1.9)
question 10-"Hand shaking improves."	**.92**	(.70 – 1.0)	.36	(.14 - .56)	**.23**	(.02 – 3.5)	1.3	(.91 – 1.9)
question 11-"Worse with hand use.."	**.92**	(.70 – 1.0)	.31	(.10 - .51)	**.27**	(.07 – 4.1)	1.4	(1.0 – 1.9)
Median sensory field 1	**.80**	(.45 – 1.2)	.53	(.30 - .75)	**.38**	(.06 – 2.3)	1.7	(.89 – 3.2)
Median sensory field 2	.60	(.17 – 1.0)	.61	(.39 - .84)	.65	(.21 – 2.0)	1.5	(.61 – 3.9)
Median sensory field 3	.50	(.01 - .99)	.68	(.48 - .89)	.73	(.26 – 2.0)	1.6	(.49 – 5.2)

Table 26 (cont'd)

MMT abductor pollicus brevis	.60 (.17 – 1.0)	**.84** (.68 – 1.0)	**.48** (.16 – 1.4)	**3.8** (1.2 – 13.4)
Tinel's A	.40 (-.03 - .83)	.55 (.33 - .77)	1.1 (.48 – 2.5)	.89 (.27 – 2.9)
Tinel's B	**.80** (.45 – 1.2)	.55 (.33 - .77)	**.36** (.06 – 2.2)	1.8 (.92 – 3.4)
upper limb tension test A	**.80** (.45 – 1.2)	.29 (.08 - .51)	.68 (.10 – 4.6)	1.1 (.66 – 1.9)
upper limb tension test B	.60 (.17 – 1.0)	.39 (.16 - .61)	1.03 (.30 – 3.5)	.98 (.44 – 2.2)
carpal compression test	.60 (.17 – 1.0)	.20 (.02 - .38)	2.0 (.50 – 8.0)	.75 (.35 – 1.6)
Phalen's test	**.80** (.45 – 1.2)	.26 (.07 - .46)	.75 (.11 – 5.1)	1.1 (.65 – 1.8)
wrist ratio (>.73)	.60 (.17 – 1.0)	**.80** (.62 - .98)	**.50** (.17 – 1.5)	**3.0** (.97 - .93)
wrist ratio (>.70)	**.80** (.45 – 1.2)	.45 (.23 - .67)	**.44** (.07 – 2.7)	1.5 (.81 – 2.6)
hand diagram	**.80** (.45 – 1.2)	.35 (.14 - .56)	.57 (.11 – 3.6)	1.2 (.71 – 2.1)
Palmar sensory latency (>3.0)	.40 (.00 - .83)	**.80** (.62 - .98)	.75 (.35 – 1.6)	**2.0** (.50 – 3.0)
Motor latency (>.50)	.40 (.00 - .83)	**.85** (.69 – 1.0)	.71 (.34 – 1.5)	**2.7** (.60 – 11.9)
Motor amplitude (<4800)	.42 (.00 - .98)	**.98** (.91 – 1.0)	.60 (.29 – 1.2)	**17.5** (.97 – 3.2)
Spontaneous activity	.60 (.17 – 1.0)	**.95** (.86 – 1.0)	**.42** (.14 – 1.2)	**11.4** (1.5 –87.5)
Functional Status Scale (>2.3)	**.80** (.45 – 1.2)	.58 (.36 - .80)	**.35** (.06 – 2.1)	1.9 (.96 – 3.8)
Fear Avoidance Behavior Questionnaire B (>31points)	.40 (.00 - .83)	**.70** (.50 - .90)	.86 (.40 – 1.9)	1.3 (.38 – 4.7)
Symptom duration a (<78 days)	**.92** (.70 – 1.1)	.44 (.21 - .68)	**.19** (.01 – 2.1)	1.6 (1.0 - 2.7)
Symptom duration (>391 days)	.50 (.01 - .99)	**.81** (.62 - 1.0)	.62 (.22 – 1.7)	**2.7** (.65 – 11.0)

The hypothesized levels of diagnostic accuracy for each clinical examination variable and the status of the hypothesis are listed in Figures 44 -45.

Variables	Hypothesis status
question 10-"Hand shaking improves.."	*Accepted*
wrist ratio	*Accepted*
hand diagram	*Accepted*
Nerve conduction and electromyography findings	*Accepted for all parameters*
Fear Avoidance Behavior Questionnaire	*Accepted, B only*
question 6-"Entire limb numb.."	*Rejected*

Figure 44 Hypothesis: Carpal tunnel syndrome clinical examination variables will demonstrate an acceptable level of predictive validity for surgical intervention (Sn or Sp ≥ .70 or LR+ ≥ 2.0 or LR- ≤ .50)

Variables	Hypothesis status
question 1-"Most bothersome symptoms.." (i (pain) and ii (numb/tingling))	*Accepted*
question 2-"Where most bothersome.." (v (hand/fingers))	*Accepted*
question 3-"Symptom behavior.." (ii (intermittent) and iii (variable))	*Accepted*
Tinel's A	*Accepted*
Carpal compression test	*Accepted*
Upper limb tension test B	*Accepted*
Visual analog scale	*Accepted*
question 4	*Accepted*
Sensation (median nerve fields 2 and 3)	*Accepted*
CCT	*Rejected*
question 1-"Most bothersome symptoms (iii (loss of feeling))	*Rejected*
question 2-"Where most bothersome.." (all response levels)	*Rejected*
question 3-"Symptom behavior.." (i (constant))	*Rejected*
question 5-"Fumbling/dropping.."	*Rejected*
question 10-"Hand shaking improves.."	*Rejected*
question 11-"Worse with hand use.."	*Rejected*
Sensation (median nerve field 1)	*Rejected*
Manual muscle test (abductor pollicus brevis)	*Rejected*
Tinel's B	*Rejected*
Phalen's test	*Rejected*
Upper limb tension test A	*Rejected*
Symptom Severity Scale	*Rejected*
Functional Status Scale	*Rejected*

Figure 45 Hypothesis: Carpal tunnel syndrome clinical examination variables will not demonstrate an acceptable level of predictive validity for surgical intervention

6.4.2.2 Patient Perception of Change Gold Standard.

Patient perception of change was not assessed for the surgical CTS group due to the small number of subjects who received surgery. Predictive validity characteristics for the same variables assessed in the surgical intervention analysis were determined using the remaining 20 CTS subjects who did not receive surgery. The revised criterion of +/- at least 3 points was used for the predictive validity analyses of improved and worsened subjects. All predictive validity characteristics of provocative tests were computed using dichotomous ratings only.

6.4.2.2.1 Improved CTS Subjects

The five subjects who improved were considered positive for the condition and the remaining 15 subjects served as the control group. The prevalence of improved subjects was 25%. Optimum cut-off points were identified using ROC curve analyses and are listed in Figures 46 – 52. Receiver operating curves did not reveal potentially useful cut-off values for the following variables and are not shown: VASworse, palmar latency, FABQ B, FABQ total, and median sensory latency. The area under the curve for these variables was ≤ .59. The predictive validity characteristics for the predictor variables are shown in Table 27. Values that met the criteria for acceptable are in bold.

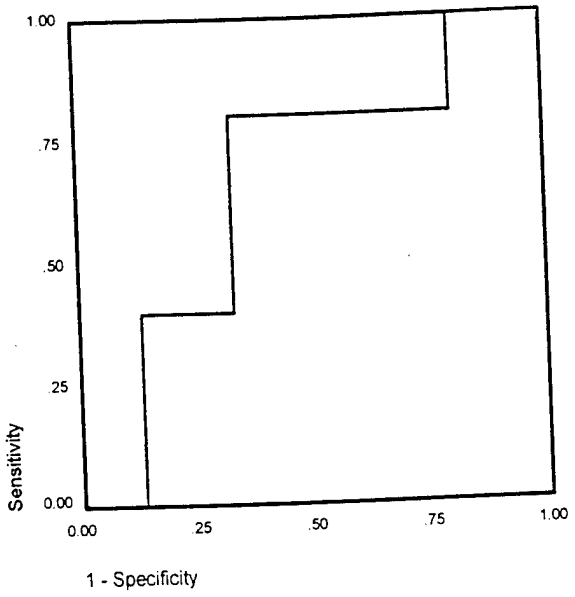


Figure 46. Receiver Operating Curve for wrist ratio: Cut-off <.70.

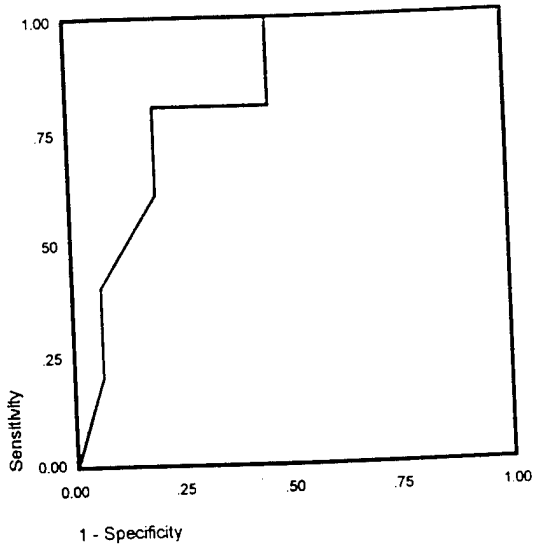


Figure 47. Receiver Operating Curve for Functional Status Scale Cut-off >1.7 points

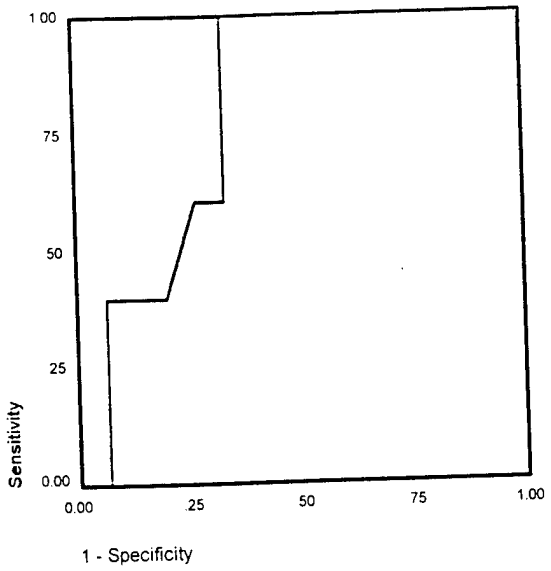


Figure 48. Receiver Operating Curve for Symptom Severity Scale: Cut-off >3.0 points.

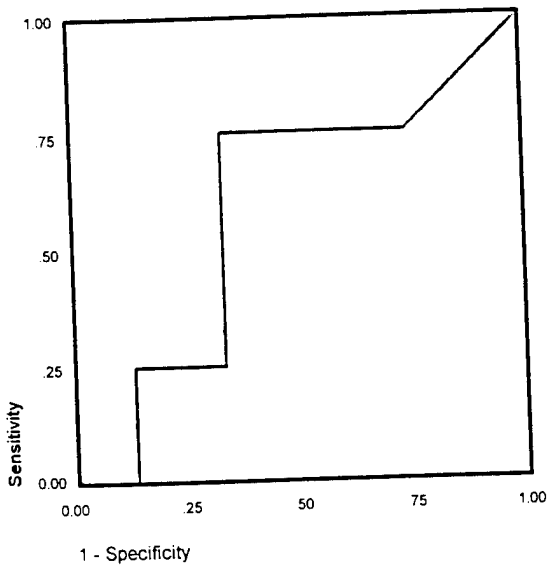


Figure 49. Receiver Operating Curve for visual analog scale (now): Cut-off >3.4 centimeters.

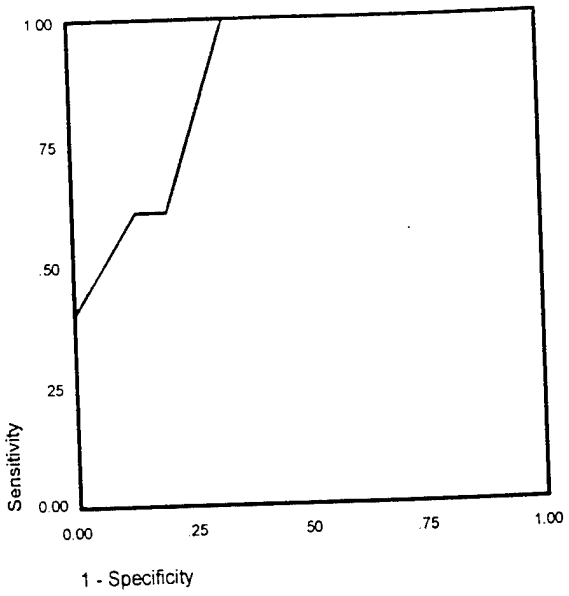


Figure 50. Receiver Operating Curve for Fear Avoidance Behavior Questionnaire A: Cut-off >54%.

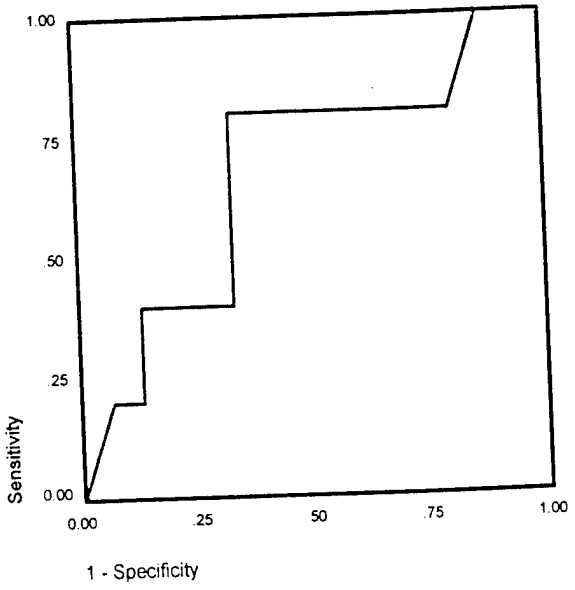


Figure 51. Receiver Operating Curve for involved median palmar sensory amplitude: Cut-off >43 microvolts

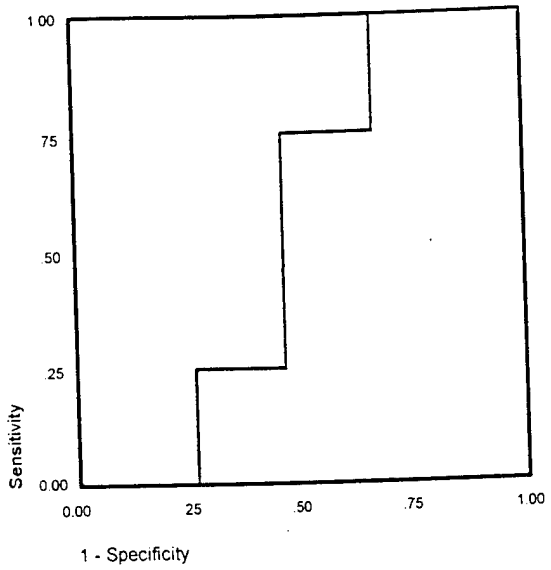


Figure 52. Receiver Operating Curve for involved median motor amplitude : Cut-off >8110 microvolts

Table 27 Predictors for improvement in non-surgical carpal tunnel syndrome subjects: Sensitivity (Sn), Specificity (Sp), Negative Likelihood Ratio (LR-), and Positive Likelihood Ratio (LR+) of clinical examination variables with 95% confidence intervals (95CI)

Variable	Sn 95CI	Sp 95CI	LR- 95CI	LR+ 95CI
question 1-"Most bothersome Sx's.				
i. Pain	.60 (.17 – 1.0)	.60 (.35 - .85)	*	1.5 (.58 – 3.9)
ii. Numb/tingling	.40 (.00 - .83)	.40 (.15 - .65)		.67 (.21 – 2.1)
iii. Loss of feeling	**	**.99** (.95 – 1.0))		**
question 2-"Where most bothersome.."				
i. Neck	.20 (.00 - .55)	**.93** (.81 – 1.0)	*	**3.0** (.23 – 39.6)
ii. Shoulder/scap.	.40 (.00 - .83)	**.80** (.60 –1.0)		2.0 (.46 – 8.8)
iii. Arm AE	**			
iv. Arm BE	**			
v. Hand or fingers	.02 (.00 - .14)	**.87** (.69 – 1.0)		.15 (.1.1 – 77.8)
question 3-"Sx. behavior.."				
i. Constant	.20 (.00 - .55)	**.80** (.60 –1.0)	*	1.0 (.13 – 7.6
ii. Intermittent	.20 (.00 - .55)	.60 (.35 - .85)		.50 (.08 – 3.2)
iii. Variable	.60 (.17 – 1.0)	.60 (.35 - .85)		1.5 (.58 – 3.9)
question 4-"Hand fat/swollen.."	**.92** (.70 – 1.0)	.66 (.42 - .89)	**.13** (.01 – 1.9)	**2.7** (1.3 – 5.5)
question 5-"Fumble and dropping.."	**.80** (.45 – 1.0)	.27 (.94 - .49)	.75 (.11 – 5.2)	1.1 (.64 – 1.9)
question 6-"Entire limb numb.."	**.80** (.45 – 1.0)	**.80** (.60 – 1.0)	**.25** (.04 - .47)	**4.0** (1.3 – 12.1)
question 8-"Night sx's. wake.."	**.80** (.45 – 1.0)	.33 (.09 - .57)	.60 (.09 – 4.0)	1.2 (.68 – 2.1)
question 10-"Hand shaking improves."	**.80** (.45 – 1.0)	.40 (.15 - .65)	.50 (.08 – 3.2)	1.3 (.73 – 2.4)
question 11-"Worse with hand use.."	**.80** (.45 – 1.0)	.33 (.09 - .57)	.60 (.09 – 4.0)	1.2 (.68 – 2.1)
Median sensory field 1	.40 (-.03 - .83)	.50 (.24 - .76)	1.2 (.29 – 2.9)	.80 (.24 – 2.6)

Table 27 (cont'd)

Median sensory field 2	.60 (.17 – 1.0)	.69 (.44 - .94)	.58 (.19 – 1.8)	**2.0** (.66 – 5.8)
Median sensory field 3	.60 (.17 – 1.0)	**.79** (.57 – 1.0)	.51 (.51 – 1.5)	**2.8** (.82 – 9.6)
MMT abductor pollicus brevis	.20 (.00 - .55)	**.86** (.67 – 1.0)	.93 (.57 – 1.5)	1.4 (.16 – 12.3)
Tinel's A	**.80** (.45 – 1.0)	.67 (.43 - .91)	**.30** (.10 – 1.8)	**2.4** (1.0 – 5.6)
Tinel's B	.60 (.17 – 1.0)	.60 (.35 - .85)	.67 (.21 – 2.1)	1.5 (.58 – 3.9)
upper limb tension test A	**.75** (.33 – 1.0)	.31 (.06 - .56)	.81 (.12 – 5.3)	1.1 (.55 – 2.1)
upper limb tension test B	**.80** (.45 – 1.0)	.46 (.19 - .73)	**.43** (.07 – 2.8)	1.5 (.76 – 2.9)
carpal compression test	**.92** (.70 – 1.0)	.28 (.06 - .50)	**.30** (.02 – 4.7)	1.3 (.96 – 1.9)
Phalen's test	**.92** (.70 – 1.0)	.33 (.12 - .61)	**.23** (.01 – 3.5)	1.5 (.92 – 2.3)
wrist ratio low (<.70)	**.80** (.45 – 1.0)	.67 (.43 - .91)	**.30** (.05 – 1.8)	**2.4** (1.0 – 5.6)
hand diagram	.60 (.17 – 1.0)	.33 (.09 - .57)	1.2 (.33 – 4.4)	.90 (.40 - 2.0)
palmar sensory amplitude (>4.3)	**.80** (.45 – 1.0)	.67 (.43 - .91)	**.30** (.05 – 1.8)	**2.4** (1.0 – 5.6)
motor amplitude (>8110)	**.80** (.45 – 1.0)	.67 (.43 - .91)	**.30** (.05 – 1.8)	**2.4** (1.0 – 5.6)
spontaneous activity	.60 (.17 - 1.0)	.67 (.43 – .90)	.60 (.19 – 1.9)	1.8 (.62 – 5.3)
Functional Status Scale (>1.7)	**.92** (.70 – 1.0)	**.85** (.63 – 1.0)	**.10** (.01 – 1.4)	**6.1** (1.4 – 27.5)
Symptom Severity Scale (>3.0)	**.92 (.70 - 1.0)**	.66 (.42 - .89)	**.13 (.01 – 1.9)**	**2.7 (1.3 – 5.5)**
Fear Avoidance Behavior Questionnaire A (>54%)	**.92** (.70 – 1.0)	.67 (.43 - .91)	**.03** (.00 – 13.8)	**2.9** (1.4 – 6.1)
visual analog scale now (>3.4)	**.75** (.33 – 1.0)	.66 (.42 - .89)	**.13** (.01– 1.9)	**2.7** (1.3 – 5.5)
age >40	**.92 (.71 – 1.0)**	.34 (.11 - .58)	**.24** (.02 – 3.0)	1.4 (.91 – 2.1)

The hypothesized level of diagnostic accuracy for each clinical examination variable and the status of the hypothesis is listed in Figures 53 - 54.

Variables	Hypothesis status
question 6-"Entire limb numb.."	*Accepted*
question 10-"Hand shaking improves.."	*Accepted*
wrist ratio	*Accepted*
Nerve conduction and electromyography findings	*Accepted*, all parameters except spontaneous activity
Fear Avoidance Behavior Questionnaire	*Accepted*, A only
hand diagram	*Rejected*

Figure 53. Hypothesis: Carpal tunnel syndrome clinical examination variables will demonstrate an acceptable level of predictive validity for improvement (Sn or Sp ≥ .70 or LR+ ≥ 2.0 or LR- ≤ .50)

Variables	Hypothesis status
question 1-"Most bothersome symptoms (i (pain) and iii (loss of feeling)),	*Accepted*
question 2-"Where most bothersome.." (iii (arm AE) and iv (arm BE))	*Accepted*
question 3-"Symptom behavior.." (ii (intermittent) and iii (variable))	*Accepted*
Tinel's B	*Accepted*
question 4-"Hand fat/swollen.."	*Rejected*
question 5-"Fumbling/dropping.."	*Rejected*
question 8-"Night symptoms wake.."	*Rejected*
question 10-"Hand shaking improves.."	*Rejected*
question 11-"Worse with hand use.."	*Rejected*
Sensation	*Rejected*
Manual muscle test (abductor pollicus brevis) 6	*Rejected*
Tinel's A	*Rejected*
Carpal compression test	*Rejected*
Phalen's test	*Rejected*
Upper limb tension test A	*Rejected*
Upper limb tension test B	*Rejected*
Visual analog scale	*Rejected*
Symptom Severity Scale	*Rejected*
Functional Status Scale	*Rejected*

Table 54. Hypothesis: Carpal tunnel syndrome clinical examination variables will not demonstrate an acceptable level of predictive validity for improvement

6.4.2.2.2 Worsened CTS Subjects

The seven subjects who worsened were considered positive for the condition and the remaining 15 subjects served as the control group. The prevalence of worsened subjects was 35%. All predictive validity characteristics of provocative tests were computed using dichotomous ratings only. Optimum cut-off points were identified using ROC curve analyses and are listed in Figures 55 – 58. Receiver operating curves did not reveal potentially useful cut-off values for the following variables and are not shown: VASnow, FABQ A, B and Total, median palmar sensory latency and amplitude, median motor latency latency and amplitude. The area under the curve for each of these variables was less than .47 except for VASnow. The VASnow variable had an area under the curve of.69, which was less than the VASworse area of .87 and therefore was not included. The predictive validity characteristics for the predictor variables are shown in Table 28.

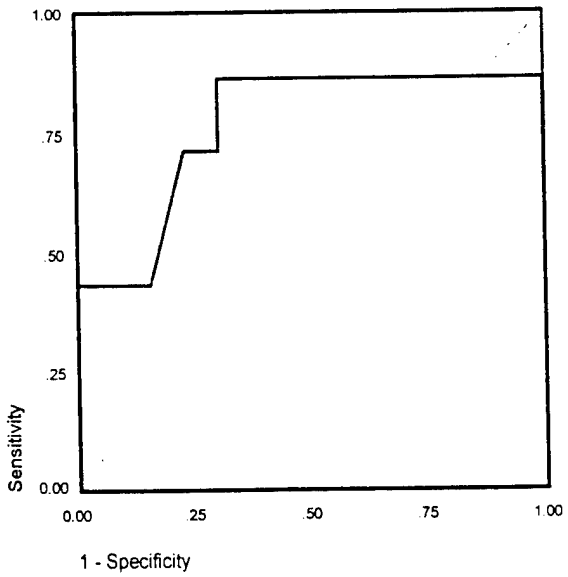


Figure 55. Receiver Operating Curve for wrist ratio : Cut-off >.70.

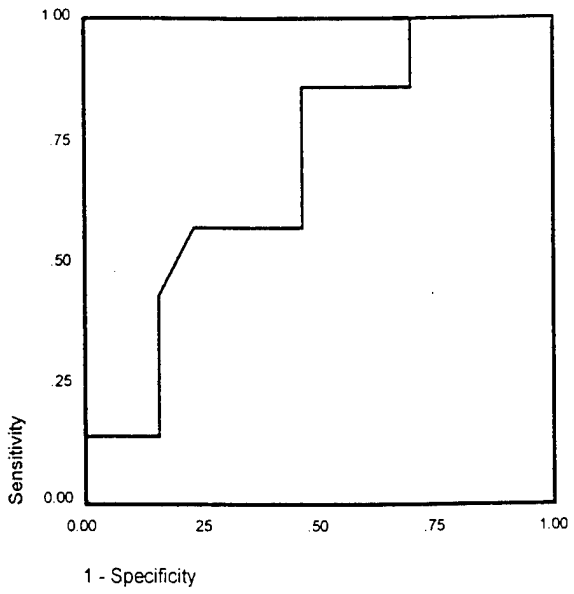


Figure 56. Receiver Operating Curve for Symptom Severity Scale : Cut-off >2.0 points

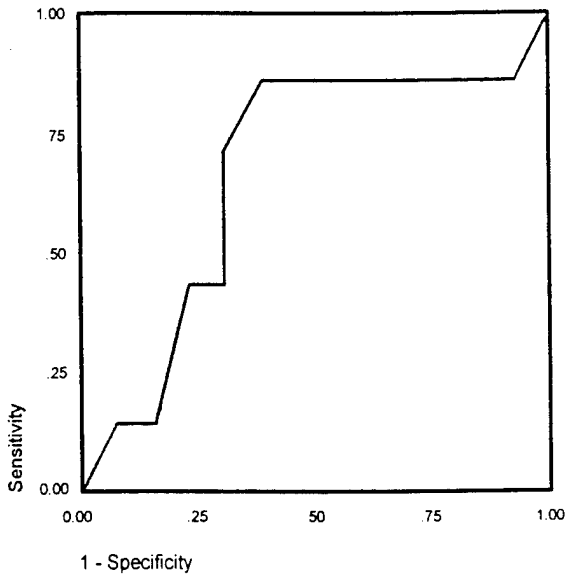


Figure 57. Receiver Operating Curve for Functional Status Scale : Cut-off >2.0 points

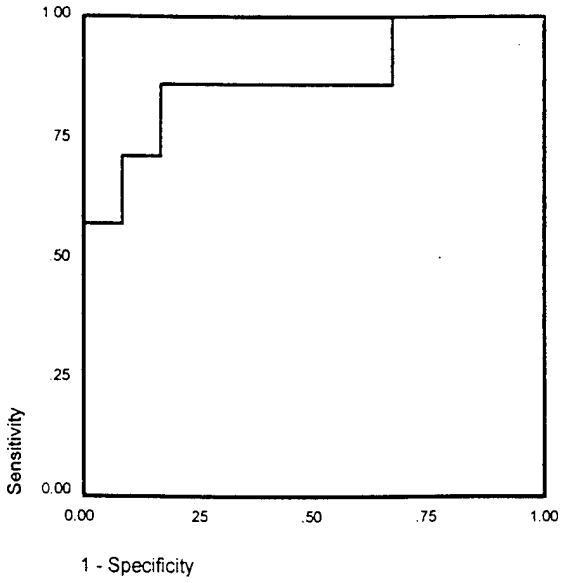


Figure 58. Receiver Operating Curve for visual analog scale (worse) : Cut-off >4.9 centimeters

Table 28 Predictors of worsening in non-surgical carpal tunnel syndrome subjects: Sensitivity (Sn), Specificity (Sp), Negative Likelihood Ratio (LR-), and Positive Likelihood Ratio (LR+) of clinical examination variables with 95% confidence intervals (95CI)

Variable	Sn 95CI	Sp 95CI	LR- 95CI	LR+ 95CI
question 1-"Most bothersome Sx's.				
i. Pain	.43 (.06 - .80)	.54 (.27 - .54)	*	.93 (.33 – 2.6)
ii. Numb/tingling	.57 (.20 - .94)	.46 (.19 - .46)		1.1 (.47 – 2.4)
iii. Loss of feeling	**			**
question 2-"Where most bothersome.."				
i. Neck	**	**.77** (.54 – 1.0)		**
ii. Shoulder/scap.	.29 (.00- .62)	**.77** (.54 – 1.0)	*	1.2 (.27 – 5.8)
iii. Arm AE	.14 (.00 - .46)	**.96** (.87 – 1.0)		**5.3** (.24 – 114.3)
iv. Arm BE	.31 (.00 - .63)	**96** (.87 – 1.0)		**8.8** (.48 – 160.5)
v. Hand or fingers	.29 (.00 - .62)	.38 (.12 - .65)		.46 (.13 – 1.6)
question 3-"Sx. behavior.."				
i. Constant	.01 (.00 - .10)	.69 (.44 - .94)	*	.05 (.00 – 22.7)
ii. Intermittent	.43 (.06 - .80)	.69 (.44 - .94)		1.4 (.43 – 4.5)
iii. Variable	.57 (.20 - .94)	.62 (.35 - .88)		1.5 (.58 – 3.8)
question 4-"Hand fat/swollen.."	.57 (.20 - .94)	.54 (.27 - .81)	.80 (.30 – 2.2)	1.2 (.52 – 3.0)
question 5-"Fumble and dropping.."	**.94** (.39 - .90)	.39 (.14 - .65)	**.16** (.01 – 2.5)	1.6 (.98 – 2.4)
question 6-"Entire limb numb.."	.29 (.00 -.62)	.62 (.35 - .88)	1.2 (.61 – 2.2)	.74 (.19 – 2.9)
question 8-"Night sx's. wake.."	**.83** (.54 – 1.0)	.43 (.17 - .69)	**.39** (.06 - 2.6)	1.5 (.82 – 2.6)
question 10-"Hand shaking improves."	**.86** (.46 - .60)	.46 (.19 - .73)	**.31** (.05 – 2.1)	1.6 (.88 – 2.9)
question 11-"Worse with hand use.."	**.71** (.38 – 1.0)	.31 (.06 - .56)	.93 (.22 – 3.9)	1.0 (.57 – 1.9)
Median sensory field 1	.50 (.10 - .90)	.54 (.27 - .81)	.93 (.36 – 2.4)	1.1 (.40 – 2.9)
Median sensory field 2	.40 (.03 - .83)	.62 (.35 - .88)	.97 (.42 – 2.3)	1.0 (.29 – 3.7)

Table 28 (cont'd)

Median sensory field 3	.17 (.13 - .46)	.62 (.35 - .88)	1.4 (.77 – 2.4)	.43 (.06 – 3.0)
MMT abductor pollicus brevis	.17 (.13 - .46)	**.85** (.65 – 1.0)	.98 (.64 – 1.5)	1.1 (.12 – 9.8)
Tinel's A	.29 (.00 - .62)	.46 (.19 - .73)	1.6 (.73 - 3.28	.53 (.15 – 1.9)
Tinel's B	.43 (.06 - .80)	.54 (.27 - .81)	1.1 (.47 – 2.4)	.93 (.33 – 2.6)
upper limb tension test A	**.80** (.45 – 1.0)	.33 (.07 - .60)	.60 (.09 – 4.1)	1.2 (.66 – 2.2)
upper limb tension test B	**.80** (.45 – 1.0)	.46 (.19 - .73)	**.43** (.07 – 2.8)	1.5 (.76 – 2.9)
carpal compression test	**.86** (.46 - .60)	.23 (.00 - .46)	.62 (.08 - 4.9)	1.3 (.67 – 2.6)
Phalen's test	**.93** (.74 – 1.0)	.39 (.14 - .38)	**.18** (.01 – 2.8)	1.5 (.96 – 2.4)
wrist ratio high (>.70)	**.86** (.60 – 1.0)	.62 (.35 - .88)	**.23** (.04 – 1.5)	**2.2** (1.1 – 4.72)
hand diagram	**.86** (.60 – 1.0)	.46 (.19 - .73)	**.31** (.05 – 2.1)	1.6 (.88 - 2.9)
Spontaneous activity	.14 (.00 - .40)	**.92** (.78 – 1.0)	.93 (.22 – 3.9)	1.9 (.14 – 25.4
Functional Status Scale (>2.0)	**.86** (.60 – 1.0)	.62 (.35 - .88)	**.23** (.04 - .50)	**2.2** (1.1 – 4.7)
Symptom Severity Scale (>2.0)	**.94** (.77 – 1.0)	.32 (.08 - .57)	**.19** (.01 – 3.2)	1.4 (.92 – 2.1)
Visual analog scale worse (>4.9)	.86 (.00 – 1.0)	**.83** (.62 – 1.0)	**.17** (.03 – 1.1)	**5.1** (1.4 – 18.9)

The hypothesized level of diagnostic accuracy for each clinical examination variable and the status of the hypotheses are listed in Figures 59 – 60.

Variables	Hypothesis status
question 10-"Hand shaking improves.."	*Accepted*
wrist ratio	*Accepted*
hand diagram	*Accepted*
Nerve conduction and electromyography findings	*Accepted:* Spontaneous activity in abductor pollicus brevis *Rejected:* all other parameters
question 6-"Entire limb numb.."	*Rejected*
Fear Avoidance Behavior Questionnaire	*Rejected*

Figure 59. Hypothesis: Carpal tunnel syndrome clinical examination variables will demonstrate an acceptable level of predictive validity for worsening (Sn or Sp ≥ .70 or LR+ ≥ 2.0 or LR- ≤ .50)

Variables	Hypothesis status
question 1-"Most bothersome symptoms.."(i (pain), ii (numb/tingling), iii (loss of feeling))	*Accepted*
question 2 (v (hands/fingers))	*Accepted*
question 3-"Symptom behavior.." (i (constant), ii (intermittent), iii (variable))	*Accepted*
question 4-"hand fat/swollen.."	*Accepted*
Sensation (all median nerve fields)	*Accepted*
Tinel's A	*Accepted*
Tinel's B	*Accepted*
questions 5-"Fumbling/dropping.."	*Rejected*
question 8-"Night symptoms wake.."	*Rejected*
question 10-"Hand shaking improves.."	*Rejected*
question 11-"Worse with hand use.."	*Rejected*
Manual muscle test (abductor pollicus brevis)	*Rejected*
Carpal compression test	*Rejected*
Phalen's test	*Rejected*
Upper limb tension test A	*Rejected*
Upper limb tension test B	*Rejected*
Visual analog scale	*Rejected*
Symptom Severity Scale	*Rejected*
Functional Status Scale	*Rejected*

Figure 60 Hypothesis: Carpal tunnel syndrome clinical examination variables will not demonstrate an acceptable level of predictive validity for worsening

In summary, Tables 29 through 31 lists the variables with acceptable predictive Likelihood ratios for surgical intervention and for change in non-surgically treated CTS subjects.

Table 29. Acceptable Likelihood ratios of surgical predictors. (LR+ ≥ 2.0 or LR- ≤ .50)

Category/Test item	Negative Likelihood Ratio	Positive Likelihood Ratio
Questions		
question 1-"Most bothersome symptoms.." (iii (loss of feeling))	--	10.5
question 5-"Fumbling/ dropping.."	.32	--
question 10-"Hand shaking improves.."	.23	--
question 11-"Worse with hand use.."	.27	--
Neurologic Examination		
median nerve field 1	.38	--
abductor pollicus brevis muscle test	.48	3.8
Self-Report		
Functional Status Scale	.35	--
Provocative Tests & Measures		
Tinel's B	.36	--
wrist ratio >.73	.50	3.0
wrist ratio >.70	.44	--
EMG/NCS (median nerve parameters; ms=milliseconds; uv= microvolts)		
palmar sensory latency (>3.0ms)	--	2.0
motor latency (>5.0ms)	--	2.7
Motor amplitude (<4800uv)	--	17.5
Spontaneous activity (abductor pollicus brevis muscle)	.42	11.4
Additional Predictors		
Symptom duration (<78 days)	.19	--
Symptom duration (>391 days)	--	2.7

Table 30. Acceptable Likelihood ratios of improved predictors. (LR+ ≥ 2.0 or LR- ≤ .50)

Category/Test item	Negative Likelihood Ratio	Positive Likelihood Ratio
Questions		
question 2-"Where most bothersome.." (i (neck))	--	3.0
question 2-"Where most bothersome.."(ii (shoulder, shoulder blade))	--	2.0
question 4-"Hand fat/swollen.."	.13	2.7
question 6-"Entire limb numb.."	.25	4.0
Neurologic Examination		
median nerve field 2		2.0
median nerve field 3		2.8
Self-Report		
Functional Status Scale (>.1.7)	.10	6.1
Symptom Severity Scale(>3.0)	.13	2.7
Fear Avoidance Behavior Questionnaire A (>54%)	.13	2.9
Visual analog scale(now) (>3.4)	.13	2.7
Provocative Tests & Measures		
Tinel's A	.30	2.4
upper limb tension test B	.43	--
Carpal compression test	.30	--
Phalen's test	.23	--
wrist ratio >.70	.30	2.4
EMG/NCS (median nerve parameters; ms=milliseconds; uv= microvolts)		
palmar sensory amplitude (>43uv)	.30	2.4
motor amplitude (>8100uv)	.30	2.4

Table 31. Acceptable Likelihood ratios of worsened predictors. (LR+ ≥ 2.0 or LR- ≤ .50)

Category/Test item	Negative Likelihood Ratio	Positive Likelihood Ratio
Questions		
question 2-"Where most bothersome.." (iii (arm AE))	--	5.3
question 2-"Where most bothersome.." (arm BE (iv))	--	8.8
question 5-"Fumbling/dropping.."	.16	--
question 8-"Night symptoms wake.."	.39	--
question 10-"Hand shaking improves.."	.31	--
Self-Report		
Functional Status Scale (>2.0)	.23	2.2
Symptom Severity Scale(>2.0)	.19	--
Visual analog scale(now) (>4.9)	.17	5.1
Hand diagram (1 or 2 or3)	.16	--
Provocative Tests & Measures		
upper limb tension test B	.43	--
Phalen's test	.18	--
wrist ratio >.70	.23	2.2

6.5 Hypothesis #4 – Test Item Cluster (TIC)

A binary logistic regression model was used solely as a variable reduction method to identify the most accurate and parsimonious TIC for all diagnostic tests and their respective dependent variable conditions.[23] Duration of symptoms was eliminated from the model because duration of symptoms was missing for 20% – 25 % of subjects in this analysis. Due to the large number of predictor variables, only variables with an LR+ point estimate ≥ 2.0 or an LR- point estimate of ≤ .50 were entered in to the regression model. This preliminary variable reduction method represents a change from the criteria originally specified due to the wide 95CI's for the test variable LR's in this study. All continous measures were entered into the model as dichotomized variables based on

previously established cut-off values. However, variables that had multiple unique-level response items (questions 1 – 3) were not transformed prior to entering the model. The Hosmer-Lemeshow (HL) summary goodness-of-fit statistic was used to assess the fit of the model to the data. The HL test is based on the distance between the observed and fitted values, is well approximated by the Pearson Chi-square distribution with df= $g - 2$ (g is the percentile-type grouping of observed and fitted values, with g usually equal to 10 groups), and provides a single, easily interpretable value which can be used to assess model-data fit. The HL tests the hypothesis that the model fits the data. Therefore, higher values of p indicate a better fit.[272pp 142 - 144] Because the purpose of the logistic regression model was strictly variable reduction, analyses of the individual residuals and additional diagnostic statistics of the fitted model were not performed.[23]

6.5.1 Diagnostic Characteristics of the CR TIC

For the CR TIC,d backward step-wise selection was used to enter variables into the logistic regression model with p values of .1 and .15 to enter and exit the model, respectively. The method of entry and liberal p-values were chose in order to prevent potentially useful variables from being excluded from the model.[273]. The following test variables were entered into the regression model as predictors for CR: ULTTA, question 7-"Symptoms keep from falling asleep..", question 9-"Neck movement improves..", Valsalva, biceps brachii MSR, Distraction, VASworse, Spurling A, question 2, MMT biceps, and MMT deltoid. After list-wise deletion, a total of 60 subjects were used in the analysis. The results of the HL test indicated the model fit the data (final p= .93). The four test variables chosen by the model and therefore considered the best CR TIC include: question 9, Valsalva, biceps brachii MSR, and Distraction. The diagnostic characteristics of the CR TIC were calculated as for other dichotomous variables using a typical 2 X 2 contingency table. When a zero cell value was encountered, .5 was added to all cell values in the table. Three different criterion levels for a positive test were established based on number of positive findings for variables in the TIC. Table 32 lists the diagnostic characteristics of the CR TIC. Values that were considered acceptable are in bold.

Table 32 Test item cluster for the diagnosis of cervical radiculopathy : question 9-"Neck movement improves..", Valsalva, Biceps brachii muscle stretch reflex, and distraction. Sensitivity (Sn), Specificity (Sp), Negative Likelihood Ratio (LR-), and Positive Likelihood Ratio (LR+) of clinical examination variables with 95% confidence intervals (95CI)

Number of any positive findings in cluster	Sn 95CI	Sp 95CI	LR- 95CI	LR+ 95CI
One	**.85** (.65 – 1.0)	.57 (.44 - .69)	**.27** (.07 - .99)	**2.0** (1.3 – 2.8)
Two	.54 (.27 - .81)	**.93** (.87 - .1.0)	**.50** (.27 - .89)	**8.1** (2.8 – 23.6)
Three	.25 (.01 - .48)	**.99** (.97 – 1.0)	.76 (.56 – 1.0)	**30.5** (1.7 – 557.5)

The hypothesis that a combination of clinical examination variables and/or patient self-report items can be found to yield acceptable levels of diagnostic accuracy for CR is accepted.

The hypothesis regarding the predictive validity of a TIC for surgical intervention and improvement in subjects CR could not be tested due to the absence of surgical intervention and the low prevalence of CR subjects.

6.5.2 CTS TIC's

Predictor variables for the CTS diagnosis TIC were entered into the regression model using the same variable entry procedure described for the CR TIC. However, all the predictive validity CTS TIC's still had a large number of variables using the initial variable reduction criteria. Therefore, the following revised preliminary variable reduction procedure and criteria were used: 1 Only variables with an LR+ of ≥ 3.0 or an LR- value of $\leq .30$ were entered in the model; 2. Variables with both LR+ and LR- values that met the previous inclusion criteria of ≥ 2.0 or $\leq .50$, respectively, were transformed to the same scale by taking the natural log of both LR's. The natural log values were then summed to a single LRprime value and variables with an LRprime value of 1.2 were included in the model. An LR prime value of 1.2 is equivalent to LR+ and LR- values of .30 and 3.4, respectively. For predictive validity TIC's, the additional predictor variables previously considered were also assessed, which included: FABQA

score, FABQB score, palmar sensory latency, motor latency; palmar sensory amplitude, motor amplitudes, and the EMG spontaneous activity rating of the abductor pollicus brevis.

6.5.2.1 Diagnostic Characteristics of the CTS diagnosis TIC

The following test variables were entered into the regression model: FSS (>2.5), question 10-"Hand shaking improves..", SSS (>3.2), question 3-"Symptom behavior..", question 5-"Fumbling/dropping..", and MMT abductor pollicus brevis. After list-wise deletion, a total of 73 subjects were used in the analysis. The results of the HL test indicated the model fit the data (final p= .77). The four test variables chosen by the model and therefore considered the best CTS TIC include: questions 3-"Symptom behavior..", and 10-"Hand shaking improves..", SSS, and MMT abductor pollicus brevis. Question 3 was dichotomized into positive (i (constant)) and negative (ii (intermittent) and iii (variable)) responses. The diagnostic characteristics of the CTS TIC were calculated as for other dichotomous variables using a typical 2 X 2 contingency table. As with the CR diagnosis TIC, three different criterion levels for a positive test were established based on number of positive findings for variables in the TIC. Table 33 lists the diagnostic characteristics of the CTS diagnosis TIC

Table 33. Test item cluster for the diagnosis of carpal tunnel syndrome : question 10-"Hand shaking improves..", question 3-"Symptom behavior..", Symptom Severity Scale (>3.2), and abductor pollicus brevis muscle test. Sensitivity (Sn), Specificity (Sp), Negative Likelihood Ratio (LR-), and Positive Likelihood Ratio (LR+) of clinical examination variables with 95% confidence intervals (95CI)

Number of any positive findings in cluster	Sn 95CI	Sp 95CI	LR- 95CI	LR+ 95CI
One	**.94** (.84 – 1.0)	.47 (.32 - .61)	**.14** (.04 - .55)	1.7 (1.3 – 2.4)
Two	.45 (.28 - .63)	**.82** (.79 - .93)	.67 (.47 - .95)	**2.5** (1.2 - 5.2)
Three	.16 (.03 - .29)	**.98** (.93 – 1.0)	.86 (.73 – 1.0)	**7.1** (.87 – 57.8)

The hypothesis that a combination of clinical examination variables and/or patient self-report items can be found to yield acceptable levels of diagnostic accuracy for CTS is accepted.

6.5.2.2 Diagnostic Characteristics of the CTS Surgery TIC

The following test variables were entered into the regression model: questions 1-"Most bothersome symptoms..", 2-"Where most bothersome..", 10-"Hand shaking improves..", 11-"Worse with hand use..", 5-"Fumbling/dropping..", MMT abductor pollicus brevis, and wrist ratio(> .73). After list-wise deletion, a total of 25 subjects were used in the analysis. The results of the HL test indicated the model fit the data (final p= .99). The four test variables chosen by the model and therefore considered the best CTS TIC include: questions 1-"Most bothersome symptoms.." and 2-"Where most bothersome..", MMT abductor pollicus brevis, and wrist ratio (>.73). Question 1 was dichotomized into positive (iii (loss of feeling)) and negative (i (pain) and ii (numb/tingling)) responses. Question 2 was also dichotomized into positive (v (hand/fingers)) and negative (i (neck), ii (shoulder/shoulder blade), iii (arm AE), and iv (arm BE)) responses. The diagnostic characteristics of the CTS TIC were calculated as for other dichotomous variables using a typical 2 X 2 contingency table. Three different criterion levels for a positive test were established based on number of positive findings for variables in the TIC. Table 34 lists the diagnostic characteristics of the CTS surgical TIC.

Table 34 Test item cluster for carpal tunnel syndrome surgical predictors: question 1-"Most bothersome symptoms..", question 2-"Where most bothersome..", MMT abductor pollicus brevis, and wrist ratio (>.73). Sensitivity (Sn), Specificity (Sp), Negative Likelihood Ratio (LR-), and Positive Likelihood Ratio (LR+) of clinical examination variables with 95% confidence intervals (95CI)

Number of any positive findings in cluster	Sn 95CI	Sp 95CI	LR- 95CI	LR+ 95CI
One	**.92** (.70 – 1.0)	.31 (.11 - .51)	**.27** (.02 – 4.1)	1.3 (.91 – 1.9)
Two	**.92** (.70 – 1.0)	**.83** (.99 -- .83)	**.10** (.01 – 1.4)	**5.5** (2.1 – 14.7)
Three	.42 (.92 - .81)	**.98** (.91 – 1.0)	.60 (.30 – 1.0)	**17.5** (.97 – 317.3)

The hypothesis that a combination of clinical examination variables and/or patient self-report items can be found to yield acceptable levels of diagnostic accuracy for CTS is accepted.

6.5.2.3 Diagnostic Characteristics of the CTS Improved TIC

After preliminary variable reduction, the following test variables met the criteria and were initially entered into the regression model: FSS (>1.7), FABQA (>54%), questions 4-"Hand fat/swollen.." and 6-"Entire limb numb..", CCT, wrist ratio (<.70), Phalen's test, TinelA, palmar sensory amplitude (>4.3), motor amplitude (>8110), question 2-"Where most bothersome..", and VASnow (>3.4). Due to the large number of variables, the co-variance matrix of the regression model could not be calculated. The model was re-calculated after one variable at a time was removed. Variables were removed from the model based on the lowest LRprime value. After a total of six iterations, a solution to the model was reached with the following variables entered into the model: FSS, FABQA, questions 4-"Hand fat/swollen.." and 6-"Entire limb numb..", CCT, and wrist ratio (<.70). No missing data were identified by list-wise deletion and all 20 subjects were used in the analysis. The results of the HL test indicated the model fit the data (final p= 1.0). The three test variables chosen by the model and therefore considered the best CTS TIC include: question 4, FABQA (>54%), and wrist ratio (<.70). The diagnostic characteristics of the CTS TIC were calculated as for other dichotomous variables using a typical 2 X 2 contingency table. Three different criterion levels for a positive test were established based on number of positive findings for variables in the TIC. Table 35 lists the diagnostic characteristics of the CTS improved TIC.

Table 35 Test item cluster for predictors of improvement of non-surgical patients with carpal tunnel syndrome : question 4-"Hand fat/swollen..", Fear Avoidance Behavior Quesionaire A (>54%), and wrist ratio (<.70). Sensitivity (Sn), Specificity (Sp), Negative Likelihood Ratio (LR-), and Positive Likelihood Ratio (LR+) of clinical examination variables with 95% confidence intervals (95CI)

Number of any positive findings in cluster	Sn 95CI	Sp 95CI	LR- 95CI	LR+ 95CI
One	**.92** (.70 – 1.0)	.34 (.11 - .34)	**.24** (.02 – 3.8)	1.4 (.91 – 2.2)
Two	**.92** (.70 – 1.0)	.66 (90 - .66)	**.13** (.01 – 1.8)	**2.7** (1.3 – 5.5)
Three	**.75** (.40 – 1.0)	**.97** (.88 – 1.0)	**.26** (.06 – 1.0)	**24.0** (1.5 – 381.9)

The hypothesis that a combination of clinical examination variables, and/or patient self-report items, and/or EMG/NCS findings can be found to yield acceptable levels of predictive validity for improved non-surgical CTS subjects is accepted.

6.5.2.4 Diagnostic Characteristics of the CTS Worsened TIC

After preliminary variable reduction, the following test variables met the criteria and were initially entered into the regression model: Phalen's test, Hand diagram, questions 2-"Where most bothersome..", 5-"Fumbling/dropping..", 4-"Hand fat/swollen..", 10-"Hand shaking improves..", 8-"Night symptoms wake..", SSS (>2.0) wrist ratio (>.70), and FSS (>2.0). Due to the large number of variables, the co-variance matrix of the regression model could not be calculated. The model was re-calculated after one variable at a time was removed. Variables were removed from the model based on the lowest LRprime value. After a total of five iterations, a solution to the model was reached with the following variables entered into the model: questions 2 and 5, hand-diagram score, wrist ratio (>.70), and SSS (>2.0). The results of the HL test indicated the model fit the data (final p= 1.0). The three test variables chosen by the model and therefore considered the best CTS TIC include: questions 2, 5 and wrist ratio (>.70). The diagnostic characteristics of the CTS TIC were calculated as for other dichotomous variables using a typical 2 X 2 contingency table. Based on the LR's for question 2-"Where most

bothersome.." for the individual predictors of worsening, the question was dichotomized into positive (iii (arm AE) and iv (arm BE)) and negative (i (neck), ii (shoulder/shoulder blade), and v (hand/fingers)) responses. Three different criterion levels for a positive test were established based on number of positive findings for variables in the TIC. Table 36 lists the diagnostic characteristics of the CTS worsened TIC.

Table 36 Test item cluster for predictors of worsened non-surgical patients with carpal tunnel syndrome : question 2-"Where most bothersome..", question 5- "Fumbling/dropping..", and wrist ratio (>.70). Sensitivity (Sn), Specificity (Sp), Negative Likelihood Ratio (LR-), and Positive Likelihood Ratio (LR+) of clinical examination variables with 95% confidence intervals (95CI)

Number of any positive findings in cluster	Sn 95CI	Sp 95CI	LR- 95CI	LR+ 95CI
One	**.94** (.77 – 1.0)	.18 (.00 - .38)	**.35** (.02 – 6.4)	1.1 (.84 – 1.6)
Two	**.94** (.77 – 1.0)	**.75** (.52 - .97)	**.08** (.01 – 1.2)	**3.8** (1.4 – 9.5)
Three	.31 (.00 – 6.3)	**.96** (.87 – 1.0)	.71 (.44 – 1.2)	**8.8** (.48 – 160.5)

The hypothesis that a combination of clinical examination variables, and/or patient self-report items, and/or EMG/NCS findings can be found to yield acceptable levels of predictive validity for worsened non-surgical CTS subjects is accepted.

7.0 SUMMARY AND CONCLUSIONS

The purpose of this study was to evaluate the efficacy of commonly used clinical examination and patient self-report measures to diagnose and predict outcome in patients with CR and CTS. The diagnosis of both conditions was based on neural impairment findings from a standardized EMG/NCS examination. Outcome was based on the patient's surgical intervention status and global rating of change scores. The study was conducted with four specific aims: 1. To determine the inter-rater reliability of the clinical examination items used for the diagnosis of CR and CTS; 2. To to determine the

diagnostic accuracy of clinical examination items and patient self-report measures used for the diagnosis of CR and CTS; To determine the predictive validity of these same clinical examination and self-report items, electrophysiological examination variables, and other selected variables for surgical intervention and a measure of patient self-reported outcome; and 4. To determine an optimum and parsimonious test cluster of the previously listed test item categories for diagnostic and predictive validity purposes.

Although planned as a multi-center study, the majority of subjects (68 of a total of 81) were enrolled at a single medical center (WHMC). The standardized electrophysiologic examination used to establish the gold standard of neural impairment was performed by a total of nine different personnel, and results from standardized clinical examination used to calculated diagnostic accuracy and predictive validity were obtained from eight different physical therapist examiners. This limits the generalizability of the study results to other medical centers with similar characteristics but does allow the results to be generalized to raters within those facilities.

The gender and age characteristics of subjects with CR and CTS in our sample is similar to that reported by other authors.[274,275] The prevalence of the two conditions (16% for CR and 38% for CTS) for our sample was consistent with our expectations. However, three subjects were diagnosed with both conditions. The prevalence of concomitant CR and CTS in our sample (18.7%) is higher than expected based on prior reports (3 – 5%).[263] The reason for this difference in prevalence for subjects with both conditions may be attributable to the standardized, rigourous work-up received by subjects in this study. Duration of symptoms of one to 6 months and one to twelve months for CR and CTS, respectively, was an eligibility criterion when into the study began. The upper limit of this criteria was eliminated for both conditions due to limited subject enrollment. Based on the median and range of duration of symptoms for both conditions, it is clear that many individuals in our study had symptoms for quite some time before they were referred for electrophysiological testing.

The limited number of CR subjects prevented a more detailed analysis of subjects based on their EMG/NCS subclassification. The diagnosis of CR was established based on spontaneous activity in one or more muscles of the standardized examination. Muscles of the standardized examination were adequate to establish the diagnosis in all cases except one. Additional muscles sampled did provide information about the possible involvement of the C5 nerve root and C8 nerve root in two cases. The C6 and C7 nerve roots were determined to be involved in all CR subjects in this study and are the two root levels most commonly involved in subjects with CR.[6,276] However, the reason for the left extremity predominance is unknown. The flexor carpi radialis H-reflex (FCR H) was prolonged in 1 subject and absent on the involved side of two subjects who were diagnosed with CR. Although others have found the FCR H to be useful for diagnosis and possibly predicting outcome in surgical patients,[70] it was not helpful for diagnosis in our study. The inability to consistently obtain this reflex may be due to technical limitions, physiologic variability, or both factors.

Approximately half of the subjects diagnosed with CTS in this study were affected bilaterally. All patients with bilateral CTS had predominance of symptoms in one extremity and it was from this extremity that clinical examination test results were obtained. Using both affected hands from the same extremity to increase sample size might have produced dependant test results and introduced a confounding factor.[16] Because false negative NCS findings may be a concern,[118] the disto-proximal ratio technique was included as part of the electrophysiological examination in order to maximize NCS sensitivity.[116,117] This technique was performed by all EMG/NCS providers without difficulty. Four subjects were diagnosed with CTS based solely on a disto-proximal ratio abnormality.

In summary, the diagnostic classification of subjects in this study is credible and potential biases were controlled for by the subject recruitment procedure and electrophysiological examination methodology used. Based on the electrophysiologic subclassification of subjects, the sample of subjects in this study are most representative of a mild to moderate spectrum of disease.

7.1 Hypothesis #1 – Reliability

Surprisingly little has been published regarding the reliability of physical examination measures for the upper extremity in general and for CR and CTS in particular.[25] Furthermore, no data has been published reporting the reliability of historical items of the clinical examination. Perhaps this is due to the fact that an accurate diagnosis can be made from information obtained from the history in the majority of cases, resulting in little if any concern about reliability.[28-30] However, the potential for miscommunication or even misleading communication between the clinician and patient is high,[277p41] and the reliability of specific questions should be established to assess their potential utility as a diagnostic test measure.

The reliability of the following types of clinical examination measures was assessed in this study: single item questions, provocative tests, conventional neurologic examination procedures, and scaled measurements of range-of-motion and wrist dimensions. Out of the 54 different clinical examination variables assessed in this study, 44 had reliability coefficients of at least Fair to Good (Kappa= .40 - .75) and ten of these were considered to be Excellent (Kappa >.75).[260] Of these ten items, five were single item questions. Many other single item questions had Kappa values approaching the Excellent level and all questions achieved at least Fair to Good levels of reliability. For eighteen of the variables, the precision of the reliability statistic was definitive and permits a confident interpretation of the reliability coefficient point estimate.[55] Ten clinical examination measures had a poor level of reliability, six of which were conventional neurological examination tests.

Low prevalence of findings appears to be an inherent problem when assessing tests of sensation and motor weakness in patients with CR.[154]. In addition to the low prevalence of the condition in this study, only certain dermatome levels and muscles may be affected even when CR is present,[21,64pp 211-244] further contributing to a low observed based rate and misleading Kappa value. The C6 and C7 nerve roots were primarily affected in this study and could explain why the MMT and sensory tests of some neurological levels were reliable and why others were not. When disease prevalence is low, subjects are

relatively homogenous and demonstrate relatively little variation between ratings.[270] While this is a valid explanation for the poor reliability of the C8 dermatome and C8 muscle tests, it would not explain the poor reliability for the C6 dermatome and MMT of the triceps brachii and flexor carpi radialis, muscles affected by C6 and C7 nerve root impairment. One study has reported Fair to Good reliability for MMT of the triceps brachii and presence of a dermatome sensory abnormality, regardless of level.[154] We also found Fair to Good reliability for detection of a any single dermatome sensory abnormality between raters but it is unknown why the reliability of MMT for the triceps brachii and flexor carpi radialis was found to be poor in this study.

Low observed base rates for MMT abnormalities were found in this study. All results from the rater pair used in the reliability analysis were one response level apart (between "F+ - G" and "N"), with the exception of three subjects who had P- to F ratings. Therefore, MMT results were dichotomized into normal or abnormal findings for the involved limb and reliability was computed based on a dichotomized test result. The difference in reliability of MMT for the abductor pollicus brevis between the entire sample of subjects and the CTS subgroup is an example of how variation can impact a reliability statistic; variation must be present in order to be explained or accounted for. When all subjects are used to compute the reliability statistic for the abductor pollicus brevis muscle, reliability is poor (Kappa= .39). When only CTS subjects are used, reliability is Fair to Good (Kappa= .65). When compared to the entire sample of subjects, the CTS subgroup had a larger proportion of subjects with weakness of the abductor pollicus brevis. The reliability of MMT in this study cannot be generalized to the more commonly used five-level MMT scheme. The five-level ordinal scale has a larger number of response levels and would likely demonstrate a lower level of reliability. Two studies have reported the reliability of MMT scoring systems of five or more levels but the results are reported as percent agreement or with a Pearson correlation coefficient and cannot be compared with our results.[155,156]

Surprisingly, both the ULTT A and B were found to have an Excellent level of reliability. The only other study to assess the reliability of this test found it to be poor (Kappa=

.35).[154] The ULTT was not operationally defined in the study by Viikari-Juntura and is the most likely explanation for the difference in their results and ours.[154] The Excellent reliability of the ULTT's found in this study is only generalizable when the limb tension tests are performed according to our operational definition. This is an important distinction because at least four variations of ULTT's have been described.[177] Our results for Spurling's, Distraction, and shoulder abduction tests are similar to those of Viikari-Juntura, although we found an Excellent level of reliability for the distraction test versus their finding of Fair to Good.[154] The Kappa statistic for Phalen's test (Kappa= .79) and Tinel's sign (Kappa A= .47 and B=.35) appear to be different (higher and lower, respectively) from the only other study that has assessed the reliability of these items.[212] Since the procedures were operationally defined, the reason for these differences is unknown. The reliability of the CCT, which was found to be Excellent, has not been previously reported.

Reliability for cervical range-of-motion measurements was lower than expected. Based on our standardized procedure for taking measurements with a bubble goniometer and a previous report,[170] we hypothesized that the reliability of cervical range-of-motion measurements would be Excellent. There are several possible reasons for our findings. We used ICC formula (2,1) to compute our reliability statistic. While Hole reported ICC values ranging from .82 to .86 (except right rotation which was .76), they did not report the ICC formula they used. Had we used formula ICC (2,k), our results would have ranged from .77 to .91, very similar to theirs. Cervical rotation measurements were taken with a Universal goniometer. The ICC's for right and left rotation equaled and exceed, respectively, those reported by Youdas et.al.,[165] who also did not report which formula they used to determine their ICC value. Based on visual assessment of stem-and-leaf plots, the distribution of measurements for right rotation appears to be smaller than for left rotation. While this may explain the disparity between the ICC's of the two measures,[255] the SEM's for the two measures do not appear to be different and indicate a similar level of precision. Precision for all cervical range-of-motion measurements, as indicated by their SEM, is less than would be desired for clinical decision making. Using

a 95CI based on our results, the true value for a cervical flexion measurement of 65 degrees could be anywhere from 56 degrees to 74 degrees, a range of nearly 20 degrees.

The ICC of both wrist measurements is less than excellent. A restriction in range and distribution scores appear to be the cause for a lower than expected level of reliability for these measurements. However, the SEM for both wrist measures is small (2.1mm), and indicates a high level of precision or small measurement error component over measurement occasions.[254]

The high levels of reliability found for most clinical examination items in this study may be due to several factors. The primary factors include: having operationally defined test procedures; raters who were briefed on test performance prior to the study; test procedures practiced by raters prior to study implementation; and a standardized scoring sheet to record results.[278] Secondary factors include the dedicated examination space set aside for physical therapist raters, allocated time for examination, and that the raters knew they were involved in a study.[277pp 43 - 47] Although patient recall may have contributed to the high reliability of single item test questions, this would have been minimized by the two day interval between the first rater and second rater questioning.

7.2 Hypothesis #2 – Test Diagnostic Accuracy

The result of any test can be interpreted as an argument to strengthen or weaken a disease hypothesis based on the information available from the patient.[54] Unlike reliability, there are no descriptive values for what constitutes an "Excellent" "Fair" or "Poor" level of diagnostic accuracy. The true value of a diagnostic test can be judged by its discriminative power, or contribution to correct decision making about the presence or absence of a disorder based on the test result.[23,54]

The criteria set forth in this study for an acceptable level of test diagnostic accuracy was based on the ability of a test to change post-test probabilities of the disorder at least 15%. Based on these criteria, variables for which Sn or Sp were $\geq .70$ were included. This was

done to prevent excluding potentially useful useful variables. Positive and negative LR's (LR+ and LR-) were the metrics used to estimate post-test probability changes because they summarize the information of both Sn and Sp and thereby represent the discriminative power of a test. Due to the limited sample size and the impact of diseased to non-diseased ratio on LR confidence intervals,[23] the 95CI of LR values for test items in this study were wide and power was not satisfactory to detect definitive findings.

7.2.1 Cervical Radiculopathy

There were 39 tests for CR and 34 met the criteria for acceptability. Of the 34 tests, all but two of these were acceptable based on Sp values alone. A much smaller number (12) had Likelihood ratios that would result in meaningful post-test probability changes. The 12 tests with acceptable Likelihood ratios were representative of the following clinical examination categories: history= 3 (numbers 2(ii), 7, and 9), conventional neurologic examination= 3 (MMT deltoid, biceps brachii, and biceps brachii MSR), provocative tests= 4 (Spurling's A, Valsalva, Distraction, and ULTT A), ROM measurements= 1 (left rotation), and self-report instruments (VASworse). Of these 12 tests, three had both LR+ and LR- values that were acceptable (two questions and left cervical rotation).

The percentage of patients with a confirmed diagnosis of CR who are offered surgical intervention can be relatively high (35%), even on the initial visit.[279]. Therefore, tests with high Sp and LR+ values are preferred. The majority of LR's in our study that were acceptable were LR+, indicating that their respective tests serve the intended purpose for the examination of CR patients and are useful for diagnosis. Many clinicians view abnormal findings of the conventional neurological examination to be the sin quo non of CR, and the finding that all neurological test items had acceptable levels of specificity appears to justify this position. However, only three of the 14 neurological test items had acceptable LR+ values that would produce meaningful changes in post-test probability. The single question items, provocative tests, and the VASworse scale had higher LR+ values than two of three conventional neurological examination items. Question 2- "Where most bothersome.."(ii (shoulder/shoulder blade)), had an acceptable LRi and appears to support the long held view that predominate scapular pain is strongly

associated with radiculopathy.[280] Question 7-"Symptoms keep from falling asleep.." had the single best LR+ value (6.5), which was also the only definitive Likelihood ratio finding, despite the limitation in power described earlier. Based on the prevalence of CR in our study (16%), a positive response to question 7 would result in a post-test probability of 55%. This means that probability of the patient having CR has now increased 39% based on the response to a single question.

The ULTT A was found to have perfect Sn in this study and resulted in a LR- of .15. Even with the low prevalence of CR in our study, a negative result on this test would result in a post-test probability of 3% and essentially rule-out CR. The ULTTA is a potentially useful screening test to determine who will undergo the time and expense of an EMG/NCS test. Cervical rotation to the involved side (only the left side in this study) was also highly Sn with an LR- comparable to the ULTT A. Because the limited number of subjects with right sided involvement, it is unknown if rotation to the involved is also Sn for subjects with right-sided CR. Interestingly, the reliability of left cervical rotation was only Good. This provides an example of why a test with less than excellent reliability must be examined closely before being dismissed as useless.[254]

Question 7-"Do your symptoms **keep** you from falling asleep at night?" was the only CR clinical examination test item with a definitive LR+ finding. One can be confident, based on the results of this study, that the true LR+ value for a "yes" response to this question lies between 2.3 to 18.0. The LR- values for the ULTT A and for left cervical rotation were not definitive and caution must exercised when interpreting the results of these two tests due to the wide 95CI.

7.2.2 Carpal Tunnel Syndrome

The treatment of CTS often follows a typical and progressive course of medication and activity modification, splinting, and surgery based on patient response to treatement.[15] Because surgical intervention for CTS is the most common surgical procedure performed in the hand,[281] may fail to relieve symptoms in a number of cases,[14] and is often based on clinical examination findings,[282] test items with high Sp and LR+ values are important to prevent unnecessary intervention. However, useful screening tests with high Sn and LR-

values are also desired since CTS is considered a work-related musculoskeletal disorder with high workers compensation costs.[281]

The diagnostic characteristics of clinical examination tests for CTS will first be discussed separately for the Mild/Moderate and Pronounced/Severe CTS subclassification groups, followed by a discussion of these same test items for the CTS group as whole. The same control group was used for both groups, therefore specificity values remained unchanged in both subclassification analyses and are not discussed.

7.2.2.1. Subclassification Group Results

For the Mild/Moderate group, 11 of the 32 tests for CTS met the acceptance criteria. Of the 11 tests, five had Likelihood ratios that would result in meaningful post-test probability changes. The five tests were representative of the following clinical examination categories: history= 2 (numbers 2-"Where most bothersome.." (iii (arm AE) and 10-"Hand shaking improves.."), conventional neurologic examination= 1 (MMT abductor pollicus brevis), and self-report instruments= 2 (SSS and FSS). For the Pronounced/Severe group, 14 tests met the criteria for acceptability. Of these 14 tests, eight had Likelihood ratios that would result in meaningful post-test probability changes. The eight tests were representative of the following clinical examination categories: history= 4 (numbers 2-"Where most bothersome.."(v (hand/fingers)), 3-"Symptom behavior..", 5-"Fumbling/dropping..", and 10-"Hand shaking improves.."), conventional neurologic examination= 3 (MMT abductor pollicus brevis, median sensory fields 1 and 3), and self-report instruments= (FSS). Of these eight tests, three had both LR+ and LR- values that were acceptable (questions 5, 10, and median sensory field 1).

A comparison of the groups based on test items with acceptable LR values show that the diagnostic characteristics of the following test items are similar for both groups: FFS, MMT abductor pollicus brevis, and question 10. Two of these items address impairment and the other relief of symptoms. Question 10, which asks "Do your symptoms improve with moving, "shaking", or positioning your hands?", has been reported in the literature as a provocative test and called the "Flick sign".[283] Although the Flick sign was not

tested, we included this question in our study because we thought it to be reflective of the Flick sign.

Test items that demonstrated potentially different diagnostic characteristics between the groups and had acceptable LR+'s include the following: For the Mild/Moderate group question 2-"Where most bothersome.." (iii (arm AE)) and the SSS; For the Pronounced/Severe group questions 2-"Where most bothersome (v (hand/fingers)), 3-"Symptom behavior (i (constant)), and 5-"Fumbling/dropping", and median sensory fields 1 and 3. Questions 2 and 3 address location and behavior of symptoms, respectively. Question 5 and median sensory fields address motor and sensory impairment, respectively. Question 8, "Do your symptoms **wake** you at night" is of interest. This question is thought to be highly diagnostic of CTS but does not appear to be useful in mild cases. The SSS is used to measure symptoms severity. From these results, it appears that severity of symptoms (SSS) and primary complaint of upper arm symptoms are useful for the diagnosis of Mild/Moderate CTS, but not for Pronounced/Severe CTS. Likewise, it appears that constant symptoms which most bothersome in the hand and/or fingers along with functional limitations of grasp and sensory impairment are useful for the diagnosis of Pronounced/Severe CTS, but not for the Mild/Moderate CTS. A larger sample size is needed to determine if these apparent differences are significant. These findings are consistent with the existence of spectrum bias.[50] Although both groups have several common symptoms and findings, subjects in the less severe group are better distinguished by diffuse sensory complaints while subjects in the more severe group are better distinguished by localized neurological impairment and functional limitations of the hand.

For the Pronounced/Severe group, three test items had both LR+ and LR- values that were acceptable (Questions 5-"Fumbling/dropping.." and 10-"Hand shaking improves.." and median sensory field 1); no test items were identified for the Mild/Moderate group.

7.2.2.2 Collective CTS Group Results

As expected, the diagnostic utility of the clinical examination test items was diminished when the two groups of subclassified CTS patients were considered together. A total of 17 of the 32 tests for CTS met the acceptance criteria. Of the 17 tests, six had LR's that would result in meaningful post-test probability changes. The six tests were representative of the following clinical examination categories: history= 3 (numbers 3-"Symptom behavior.." (i (constant)), 5-"Fumbling/dropping..", and 10-"Shaking hand improves.."), conventional neurologic examination= 1 (MMT abductor pollicus brevis), and self-report instruments= 2 (SSS and FSS). In addition to the lower number of LR's found acceptable compared to CR, the discriminative power of these Likelihood ratios was reduced when the two groups were considered together. The exception to this was question number 10. The acceptable LR+ and LR- values for question 10 remained nearly the same in the subgroup and total group analyses.

The single best LR+ value was the FSS with a cut-off value of 2.5 (LR+= 3.1); the remainder of test items had LR+ values of 2.5 or less. Based on the prevalence of CTS in our study (38%), an FSS score of >2.5 would result in a post-test probability of 66%. This means that probability of the patient having CTS has now increased 28% based on the score of this self-report instrument. Question number 10-"Hand shaking improves.." is of particular interest because of its stability across group analyses for both LR+ and LR- values. Calculating changes in probability based on clinical examination test results, if a subject responds to question 10 with a "no", then the probability of CTS being present has diminished to 20%. Unlike ULTT A for CR, the post-test probability generated by a negative response to question 10-"Hand shaking improves.." is not satisfactory to warrant its consideration as a screening test to rule-out CTS and eliminate the need for an EMG/NCS examination.

In this discussion of useful clinical examination tests, Tinel's sign, the CCT, and Phalen's test are conspicuous by their absence. Neither of these tests had accepatable Likelihood ratios in any of the analyses. However, the Sn of the CCT was .70 in the Mild/Moderate group and .80 in the Pronounced/Severe group. Due to lack of power in the subgroup

analyses, it is unknown if this increased Sn value is significant. However, based on the 95CI in the total group analysis, Sp for both the Phalen's test and the CCT was considered definitively unacceptable. Sensitivity of Tinel's sign and Sp for the hand diagram and wrist ratio were also definitively unacceptable. At least 19 studies of Phalen's and Tinel's sign have been conducted. The earlier review of these studies in Table 4 and Figure 9 listed the wide range of Sn and Sp values reported for these tests. The most methodologically sound of these studies was by De Krom et. al. who also found lower Sn and Sp values for Phalen's test and Tinel's sign.[160] The CCT has not been studied as extensively as Phalen's test and Tinel's sign but has a similar wide range of reported diagnostic accuracy values derived from methodologically deficient studies.

Although the diagnostic accuracy values derived from the initial retrospective report of the hand diagram were promising,[223] subsequent prospective investigations have reported much lower values of Sn and Sp that are similar to the values we found for Sn[16,152]. However, our Sp values appear to be much lower (.43 vs. .71) and may be due to our control group which consisted of CR patients and patients with symptoms similar to CR and CTS. The Sn and Sp values for question 5 are similar to the results of Katz et. al.[16]. Our Sp and Sn values for the wrist ratio are consistent with the findings reported by Kuhlman and Hennessey.[172]

7.2.3 Summary of Diagnostic Accuracy Findings

Any interpretation of the utility of individual test items of the clinical examination reported in this study for the diagnosis of CR or CTS must be done in the context of our neural impairment gold standard. Although no gold standard is perfect,[40] we considered neural impairment based on NCS/EMG findings to be the optimum reference criterion for the purposes of our study. The arguments regarding false positive and negative findings related to NCS procedures have been discussed in section 2.3.2 and will not be repeated here.

Another facet related to gold standard selection is our liberal criteria for a positive EMG/NCS test for both CR and CTS. It is possible that some subjects were classified as

positive for disease based on artifact and anomalous or irrelevant findings. Including a third group of asymptomatic subjects would have allowed us to assess the potential impact our liberal diagnositic criteria had on subject classification. However, performing the standardized EMG/NCS examination on entirely asymptomatic subjects would be difficult to justify both ethically and monetarily, especially in light of findings from recent studies using this design.[213,284] Both Szabo et.al.[213] as well as Gerr and Letz[284] found that using asymptomatic subjects as a control group produced falsely optimistic estimates of diagnotic accuracy for the diagnostic tests being considered. Although liberal, our criteria were based on peer-reviewed published literature.[75,113] However, using a more conservative gold standard criterion has merit since surgical decisions are often based on EMG/NCS test results.[64pp.111-160, 211-244]

The utility of the clinical examination was better for the diagnosis of CR than for CTS based on the number of acceptable Likelihood ratios, their discriminative power, and the change in post-test probability for a single test item of the clinical examination. This may be the result of spectrum bias as observed in the subclassification analyses. Identifying characteristics of subjects most probable to be classified as Pronounced/Severe[50] would increase the diagnostic utility found for several test items in this study. The impact of spectrum bias on diagnostic tests for CR is unknown and was not assessed.

The utility of single item test questions for both conditions is remarkable when compared to other categories of clinical examination items and is consistent with previous reports regarding the diagnostic power of a patient's condition history and symptoms.[28-30] However, no single test item resulted in a post-test probability greater than 66% for either condition. If maximum specificity is desired for these two conditions, more definitive test procedures must be performed. Similarly, the test item with the best single LR- for CTS did not result in adequate post-test probabilities to be considered useful as a screening tool. The ULTT A appears to be a potentially excellent and useful screening test for CR but further study with a larger sample is required to establish a definitive LR- value.

7.3 Hypothesis #3 – Test Predictive Validity

The use of patient outcome as a gold standard attempts to define the properties of a diagnostic test at the highest level of efficacy.[40] The use of an outcome reference criterion, or gold standard, in conjunction with diagnostic test methodology and indices has been used to measure the responsiveness of HSAM's[285] and more recently to predict pain relief from splinting intervention in subjects with CTS.[286] Surgical intervention status was used as one outcome criterion measure and change based on a GRCS as the other outcome criterion measure. Clinical examination items for CR were not assessed using these outcome criteria but both outcome gold standards were assessed in patients with CTS. The results of all three predictive validity analyses for CTS patients must be considered preliminary due to the small sample size, low prevalence of surgical intervention and patients classified as changed, and incomplete follow-up of 81%.

7.3.1 Surgical Intervention Gold Standard

Surgical intervention is not based solely on the decision of the surgeon and the patient; both individuals have significant input into the decision to operate. Even though the decision to intervene surgically is probably multi-factorial in nature, it is possible that several overriding factors influencing this decision can be identified. Because all patients with positive EMG/NCS test results could be considered potential surgical candidates, test with a high specificity are desired. Positively identifying which patients will receive surgery based on the results of clinical examination findings and other predictor variables would allow a closer examination and further investigation of those variables.

The use of EMG/NCS test results as test variables could be considered a form of inclusion bias. However, not all surgeons base surgical intervention on EMG/NCS test results[282] and it is unknown which EMG/NCS test items, if any, are strong predictors of surgical intervention. Therefore, we included several median nerve EMG/NCS parameters as predictor variables along with age, duration of symptoms, and FABQ scores.

The predictive validity of 32 clinical examination test items was evaluated in addition to seven additional predictor variables. Twenty of the test items and all of the additional predictor variables met the criteria for acceptability for a total of 27 items. There were 17 items that had acceptable Likelihood ratios, 11 were clinical examination items and 6 were additional predictor variables. Acceptable Likelihood ratios were representative of the following clinical examination categories: history= 5 (numbers 1-"Most bothersome symptoms (iii (loss of feeling)), 5-"Fumbling/dropping..", 10-"Hand shaking improves..", and 11-"Worse with hand use.."), conventional neurologic examination= 2 (median sensory field 1 and MMT abductor pollicus brevis), provocative tests= 1 (Tinel's B), scaled measurement= 2 (wrist ratio = >.70 and >.73), self-report instruments=1 (FSS >2.3). The only additional predictor variables that did not result in an acceptable Likelihood ratio were the FABQ A and B scales. Of these 17 variables, three had both LR+ and LR- values that were acceptable (MMT abductor pollicus brevis, wrist ratio >.73, and spontaneous EMG activity in the abductor pollicus brevis).

The abductor pollicus brevis motor amplitude (<4800uv) and presence of spontaneous activity in the EMG produced the two largest LR+ (17.5 and 11.4, respectively). The next largest LR+ (10.5) was question 1-"Most bothersome symptoms.." (iii), indicating that the patient's most bothersome symptom was loss of feeling. The remaining LR+ values were 3.5 or less. The two variables with the largest LR+ are related, so only one should be used to estimate the post-test change in probability. Based on the pre-test probability of surgery for the 25 subjects for who returns were available (20%), a subject with a motor amplitude of <4800uv now has an 81% chance of having surgery; a sizable 60% increase in probability.

A number of test items items had small LR- values indicating that unless these variables are present, surgical intervention is not likely. The smallest of these were all questions (numbers 5-"Fumbling/dropping..", 10-"Hand shaking improves..", and 11-"Worse with hand use..") with LR- values ranging from .04 to .08. Duration of symptoms less than 78 days and age < 48 years also had a low LR- value (.19 and .14, respectively). These data indicate that surgical intervention will not be performed if the subject does not have some

component of symptoms that are most bothersome in the hand, complaints of functional limitations, is able to affect symptoms with hand movement, and if the symptoms have not been present longer than 78 days. It also indicates that surgery is much less likely to be performed on older subjects during this short-term follow-up period. Using the single best test item LR- to calculate post-test odds, a negative response to question 10 diminishes the probability of surgery diminishes from 20% to 5%.

These data suggest the EMG/NCS parameters related to the loss of motor units are the strongest predictors for surgery, which is expected because patients in our study were referred from surgeons who utilize EMG/NCS testing for the management of patients with CTS. However, the number of single question items with highly predictive Sn and LR- values was not expected. Surgical intervention for the treatment of CTS appears to be based on several facets, which include impairment, symptoms, and functional limitations. These results may be useful as a starting point for further quantifying specific items that represent these facets and how they may be used to predict which CTS patient will be treated surgically. This type of prognostic information would be useful to clinicians treating CTS patients conservatively for knowing when to terminate treatment, and for surgeons in order to analyze their treatment decisions.

7.3.2 Change Gold Standard

Because only a few of the 20 non-surgical CTS groups experienced change in either direction (improved or worsened), we lowered to threshold of change from $\geq$ +/- four or more points to $\geq$ +/- three or more points on the GRCS. We justified our change in criteria based on our interest in the patients self-reported perception of change and that our original criteria may have been too stringent to detect meaningful change in a six-week time period. The results of the predictive validity analyses and the descriptive results from the HSAM's of the patients who considered themselves worsened coincide with what would be expected in a worsened patient. Unfortunately, the predictive validity results for the improved group were often paradoxical and somewhat difficult to interpret. The predictive validity results for patients who considered themselves worsened will be discussed first, followed by the results of the "improved" group.

7.3.2.1 Worsened Patients

For the identification of subjects who will worsen, tests with good screening properties are desirable to maximize identification of subjects with a potentially poor or adverse outcome.[287] For the subjects considered worsened, the predictive validity of 30 clinical examination tests items was evaluated in addition to one additional predictor variable. Age, duration of symptoms, EMG/NCS tests items (except spontaneous activity in the abductor pollicus brevis muscle), and the FABQ questionnaire were eliminated from further consideration as predictive variables by the ROC curve analysis. Seventeen of the test items met the criteria for acceptability and 12 of these had acceptable Likelihood ratios. Acceptable Likelihood ratios were representative of the following clinical examination categories: history= 5 (numbers 2-"Where most bothersome.." (iii (arm AE) and iv (arm BE)), 5-"Fumbling/dropping..", 8-"Night symptoms wake..", and 10-), provocative tests= 2 (Phalen's test and ULTT B), scaled measurement= 1 (wrist ratio = >.70), and self-report instruments= 4 (FSS >2.0, SSS >2.0, VAS >4.9, and hand diagram (1-3= positive test)). There were no conventional neurologic examination test items that were useful predictors of worsening.

Several tests had predictive validity properties useful for screening. Question 5- "Fumbling/dropping.." and the hand diagram (1-3 positive) both had an LR- value of .16. Question 5 also had a similar LR- value of .16 for predicting surgical intervention. If the hand diagram or question 5 is used for screening which patients will worsen, the pretest probability of a patient in this sample changes from 35% to 9% when a negative test result is obtained (i.e., pain drawing is not consistent with median nerve impairment). Other items with good screening properties represented the constructs of a irritability (VASnow, SSS, Phalen's, ULTT B), functional limitation (question 5 and FSS) and possible anatomical predisposition to the condition (wrist ratio). Overall, the predictive items most powerful for subjects who worsen indicate intense forearm pain and complaints of functional limitations consistent with median nerve impairment.

Test items diagnostic of worsening were pain in the forearm or upper arm and a high VASnow rating. A positive response to question number 2-"Where most bothersome.."

(iv (arm BE)), would increase the probability of a patient in this study worsening from 35% to 87%. A test that yields this information could be useful for clinicians for the purposes of prognosis and decisions regarding continued or more aggressive conservative intervention and deciding when conservative treatment options have been exhausted.

7.3.2.2 "Improved" patients

Descriptive results from the HSAM's of the patients who considered themselves improved do not coincide with what would be expected in an improved patient. There are a couple of possible reasons for this. The first reason is that these patients are not actually improved. An improvement of $\geq$ +/- 3 points on the GRCS may be only a reflection of a patients desire to please the investigator[277p41] or measurement error. A comparison of the FSS and SSS scores of patients considered improved and not improved reveals that the improved groups scores are numerically higher and at least no different. One study reported that only 13% of conservatively treated CTS patients were completely resolved at one year follow-up.[15] If these results were generalized to our patients, few if any patients would be improved given the short-term follow-up in our study and lowering the criteria for improved status was not valid. A second reason could be that this group of patients actually consider themselves to be improved despite continued symptoms and impairment. It is well known that in many instances symptoms, impairment, and functional limitations are not well correlated.[131,135] It is unknown why the GRCS did not perform similarly when used to determine patients with a worsened status. It is possible that the scale has different measurement properties for opposite ends of the scale.[288] Based on the information available, the patients classified as "improved" should be considered to be unchanged at best and still experiencing a considerable proportion of their original symptoms. The following predictive validity of this group will now be discussed from this perspective.

Positively identifying patients who improved would serve to prevent further unnecessary testing and treatment. However, in this group of patients with persistent symptoms, a test with good screening properties would be desirable for the same reasons given for detecting patients whose status will worsen. Thirty-one clinical examination tests items

was evaluated in addition to the following five additional predictor variables: Age, FABQ A, median motor and sensory amplitudes, and spontaneous activity in the abductor pollicus brevis.

Twenty-six of the test items met the criteria for acceptability and 18 of these had acceptable Likelhood ratios. Acceptable Likelihood ratios were representative of the following clinical examination categories: history= 5 (numbers 2-"Where most bothersome.." (i (neck) and ii (shoulder/shoulderblade)), 4-"Hand fat/swollen..", 6-"Entire limb numb..", and 10-"Hand shaking improves.."), conventional neurologic examination test items=2 (median sensory fields 1 and 2), provocative tests= 4 (Phalen's test, CCT, Tinel's A, and ULTT B), scaled measurement= 1 (wrist ratio = <.70), and self-report instruments= 4 (FSS >1.7, SSS >3.0, VAS >3.4, and FABQ >54%), EMG/NCS variables= 2 (median palmar amplitude >4.3uv and median motor amplitude >8110uv). Age of >40 also had an acceptable LR- value.

The SSS and the FSS instruments both had low LR-values for this group as well as the group of CTS patients whose status worsened. Other variables with low LR- values common to both groups include the Phalen's test and the ULTT B. The variable with the most useful LR- value not common to both groups was the FABQ A score. An FABQ A score of less than 54% drops the pretest probability of being in this group to a post-test probability of 3%. Question 6 was another test items thought to represent a psychological construct that had a similarly low LR- value. The reduction of relatively high levels of initial psychologic distress may account for why patients perceive themselves to be improved despite persistant symptoms and functional limitations.

The many similar test items with high predictive values for this group and the CTS worsened group are consistent with the conclusion that the "improved" group really represents patients who experience a reducition in their initial high level of psychologic distress but have little change in symptoms and function. Based on the test items, it appears that predictors for subjects who consider themselves improved are characterized by symptom predominance (SSS and VASnow) of an axial location (question 2-"Where

most bothersome.." (i (neck) and ii (shoulder/shoulder blade)), irritability (provocative testing), and psychologic distress (FABQ A, Question 6-"Entire limb numb..") but limited axonal loss. Indeed, testing "normal" for median motor and sensory amplitudes both had LR+ and LR- values that were acceptable. The wrist ratio indicates that anatomical predisposition may be useful predictive factor.

7.3.3 Predictive Validity Summary

The preliminary findings for predictors of surgical intervention and worsened CTS subjects both had face validity and may be helpful for future research. Interpretation of the subjects predicted to be improved was more difficult; most likely these subjects were unchanged and still symptomatic. Predictor test items unique to this group when compared with the worsened group are items that relate to psychological constructs and axonal injury markers; the reason for the former observation is unclear. Test items related to more severe functional limitation seemed to be more predominating for CTS worsened subjects. For all three outcome gold standards, the wrist ratio appears to be a useful predictor variable. A narrow carpal canal (ratio >.70) is predictive of subjects with a worsened status or subjects treated surgically. An acceptably wide carpal canal (ratio <.70) is predictive of subjects in the "improved" or unchanged group. Further work is needed to clarify this relationship and determine more precise diagnostic accuracy characteristics.

7.4 Hypothesis #4 – Test Item Cluster

How does one choose to incorporate the numerous results obtained from the clinical examination? Just a cursory glance at the tables listing the diagnostic accuracy values for tests related to CR or CTS gives some indication that making an intelligent decision about which test results to use can be a daunting task. Knowingly or unknowingly, this decision is made every time a clinician treats a patient with CR or CTS. Some test results may be highly relevant while others will have very little or no utility. An additional problem is that many of these tests are likely to be conditionally dependant; that is the probability of the outcome of one test is affected by the outcome of others.[23] The use of

multiple, conditionally dependent tests can result in inaccurate diagnosis and lead to further inappropriate testing and treatment.

Often, attempts to combine individual test items into a single, powerful test item cluster appear to have been a mere afterthought in an effort to improve the mediocre predictive power of single item test results.[21] Other investigators have performed fairly sophisticated multivariate analysis for the purpose of identifying an optimum predictive model, but only a few common examination test items and self-report instrument were used and the results were reported as Odds Ratios.[16] Presently, no systematic attempt has been made to identify which of the many clinical examination variables are most useful for the diagnosis of CR and CTS and quantify their utility; this was the purpose of our study. All variables thought to be potentially useful for the diagnosis of the two conditions were considered in our study, including data typically thought to be "soft".[289]

7.4.1 Variable Selection Methodolgy

Following a preliminary variable reduction procedure based on diagnostic utility, a logistic regression procedure was used for the purpose identifying a parsimonious single TIC for the conditions of interest.[23] This approach minimizes bias, minimizes conditional dependence,[23] and identifies variables that may have marginal individual predictive power but make unique contributions to a predictive model. The TIC identified by the model was then used to compute familiar and useful indexes of diagnostic accuracy.

7.4.2 Diagnostic Accuracy TIC's

The CR TIC consisted of question 9-"Neck movement improves..", Valsalva, biceps brachii MSR, and Distraction. The CTS TIC consisted of questions 10-"Hand shaking improves.." and 3-"Symptom behavior..", the SSS, and MMT of the abductor pollicus brevis. As expected, the CR and CTS TIC's yielded LR+ values that were much greater than any of the individual test items alone; over four times larger for the former condition and over twice as large for the later. In this sample with a CR prevalence of 16%, if the single best test item is used to diagnose CR (question 7-"Symptoms keep from sleep..", LR+= 6.5), a positive response results in a post-test probability of only 55% and the

diagnosis of CR is now just a little better than chance. However, a positive response to the CR TIC using a criterion of any three abnormal items (LR+= 30.5), we obtain a post-test probability of 85% and the confidence of a definitive diagnosis has greatly increased. Using a similar example, if the single best test item is used to diagnose CTS (FSS >2.5, LR+= 3.1), a positive response results in a post-test probability of 66%. Sixty-six percent is better than the pretest probability in this sample of 38%, but still not sufficient to establish a diagnosis. If a positive response to the CTS TIC using a criterion of any three abnormal items is used (LR+= 7.1), we have now increased our post-test probability to 81%.

When a criterion of any single abnormal finding was used, the CTS TIC resulted in a very low LR- value (.14), which is useful for screening purposes. This value is over 2.5 times smaller than the LR- value of any single test. This is extremely helpful because only two individual CTS test items had LR- values that met the criteria of acceptability. A LR- value of this magnitude is powerful and results in a post-test probability of 8% when a criterion of any one finding is taken as abnormal CTS TIC and the CTS TIC was negative. A criterion of any single abnormal finding for the CR TIC resulted in an LR- of .27, which is not as powerful as the LR- associated with the ULTT A alone (LR-= .15).

The CR TIC findings raise a couple of important observations. The first is that neither of the test items with the single best LR+ and LR- values is included in the TIC. This is because the logistic regression model attempts to maximize correct classification using a predefined probability level (usually .50), without regard for direction of error (i.e. false positive or false negatives). The inclusion of apparently marginal or less powerful variables may occur in order to maximize correct classification.[272p 117] This leads to the second observation: the diagnostic characteristics of an individual test item may be sufficiently powerful by itself for diagnosis or screening. This is unusual but can occur[277p 128] The ULTT A had an LR- value of .15 in this study. The post-test probability that results when the test is negative is 3% which many would consider sufficient for screening purposes given the natural history of CR. When used for screening purposes, a negative CR TIC would result in a posttest probability of 5%.

The individual test items of the CR TIC were no surprise, and are consistent with clinical experience and what has been proposed to be useful for the diagnosis of CR.[34] This not the case for the CTS TIC. With the exception of the abductor pollicus brevis MMT, many test items thought to be predictive of CTS were not included in the TIC. Items absent were not only the traditional provocative tests (Phalnen's, CCT, Tinel's sign), but the single item question related to waking night pain. This lack of utility is not explained by spectrum bias, as none of these tests were helpful in the subclassified groups of CTS patients. Given the wide range of diagnostic characteristic reported for these test items and poor methodology discussed previously, this is really no surprise. Other investigators have also found waking night pain to have no utility for the diagnosis of CTS.[16] The wrist ratio was not helpful for diagnosis when considered by itself but was useful when considered in conjunction with other test items. This test item is of particular interest because it was ubiquitous throughout all the CTS TIC's.

7.4.3 Predictive Validity TIC's

As for the single item predictive validity analyses, the predictive TIC's in this study must be considered preliminary due to the extremely small sample size and prevalence of the condition of interest. Even so, the surgery and worsened TIC's have face validity that lends credibility to the results.[290pp 1 - 5, 141 - 166] For example, subjects whose predominate complaint is loss of feeling located in the hand and/or fingers and have a weak abductor pollicus brevis are usually considered likely surgical candidates.[149] An excessively small wrist diameter (>.73) was also identified as predictive. Anatomical predisposition to median nerve compromise within the carpal tunnel could certainly contribute to persistent signs and symptoms that are unlikely to resolve spontaneously and that are unresponsive to conservative intervention. A criterion of three positive items increases the probability of surgery from 20% to 81% during the short-term follow-up period of this study. When the criterion of two positive items is used and the TIC is negative, the chance of surgery is reduced to only 2%. The diagnostic properties of this TIC can be influenced by many facets and could change when used with different population and different surgeons.

The three items of the CTS worsened TIC also have face validity and suggests a physical basis for a worsened status in these subjects. Pain that is most bothersome in the arm and forearm of CTS patients is not uncommon and is considered a sign of increased irritation.[168] Question 5-"Fumbling/dropping.." indicates patients have a significant degree of functional limitation and possible loss of sensory/motor function. For the worsened TIC, there appears to be an anatomical predisposition for subjects who progress to a worsened status (wrist ratio >.70). Using a criterion of two items for a positive test, the probability of a non-surgical subject becoming worsened in this sample changes from 35% to 4% when the TIC is negative. Based on this TIC, subjects without a 2-item combination of either arm/forearm pain, complaints of clumsy hands or dropping things, or an abnormal wrist ratio are very unlikely to progress to a worsened status over the short-term follow-up period in this study. If the criterion for positive is all three items, then a patient with a positive test has an 83% chance of becoming worsened. Patients considered to be at risk for worsening may merit closer monitoring and the decision for earlier aggressive therapy might be considered.

Because subjects in the "improved" group were considered to be unchanged subjects with persistent symptoms, no pre and post-test probability examples are given. The inclusion of the FABQ A may reflect some degree of psychological distress in these patients which was expected to be present in patients with a worsened status. However, this relationship could change with a longer follow-up period. Question 4-"Hand fat/swollen.." represents sensory disturbance as being predictive as do self-report scales of symptoms and function. Once again, the wrist ratio (<.70) was a useful predictive test in this group and coincides with the abnormal wrist ratio values that were predictive of the surgical and worsened subjects.

7.4.4 TIC Summary

In this study, an optimum cluster of test items had better diagnostic power than any single test item alone for the CR and CTS conditions. The exception was the ULTT A when used for screening purposes. Our method of adjustment for zero-cell findings in conjunction with our sample size resulted in conservative Likelihood ratio estimates for

the CR TIC and the predictive validity TIC's. Depending on the criterion used to determine a positive test, these TIC's resulted in perfect (1.0) Sn, Sp, or both. The sample size of this study produced wide 95CI's and resulted in only two definitive LR findings. Further testing with a larger sample size is required to know if the individual test items selected for the diagnostic TIC's are most predictive and to increase the precision of Likelihood ratio point estimates. The predictive validity TIC's in this study can provide a basis for the identification of tests and TIC's that are useful for the prognosis of treatment and outcome in patients with CTS.

7.5 Summary

This is the first study to concomitantly compare the diagnostic characteristics of clinical examination items and self-report instruments for the diagnosis of CR and CTS, two conditions that may produce similar signs and symptoms. These two conditions can cause differential diagnostic dilemmas.

This study has several strengths. Including subjects with one condition in the control group of the other effectively challenged the diagnostic characteristics of the tests that were considered. Bias was controlled for in several ways. First, a common, independent gold standard representing a construct of interest (neural impairment) was applied uniformly to every patient in the study to prevent work-up and inclusion bias. Second, the EMG/NCS provider and the physical therapy raters were blinded to each other's results to prevent test review and diagnostic bias. Third, subclassification of CTS subjects permitted an assessment of spectrum bias its effect on the diagnostic characteristics of test used for the diagnosis of CTS. The methodological rigor of this study increases the internal validity of our results. The number of different raters that contributed data from which diagnostic indices were calculated increases the generalizabilty of our study's findings. Assessing the reliability of tests used in our study permitted a closer examination of tests that performed poorly and did not exclude test with Poor or Fair reliability from consideration. All clinical examination procedures in this study were operationally defined

There are also several limitations of our study that must be acknowledged. The first is small number of subjects with CR that were included in the study. A larger number of CR subjects would most likely increase the variability of findings and could diminish the diagnostic characteristics we found for the test items in this study. Based on the severity of EMG/NCS findings, most CR subjects were considered to be mild. However, severe cases usually increase, not decrease the diagnostic characteristic of a test.[50] In addition, our findings were only for subjects with involvement of the C6-C7 nerve root. Second, the sample size in this study did not permit precise estimates for most diagnostic indices. Although several Sn and Sp values were considered definitively acceptable, only one individual CR test item and two TIC's (CR and surgery) had definitive LR+ values. None of these LR+ values would produce large changes in post-test probabilities based on the lower limit of the 95CI. Third, the use of an EMG/NCS gold standard may have prevented identification of subjects who had CR and CTS. No gold standard is perfect[40] and we chose to maximize specificity since diagnosis and not screening was the primary purpose of our clinical examination test items. This was reflected in the ratio of LR+ to LR- findings that were considered acceptable. Use of a different gold standard may result in different diagnostic characteristics for the test items in this study. Fourth, our outcome gold standard of surgery and change may not be useful or the best gold standards for assessing intervention and change, respectively. It is well known health care practices may vary by geographic region[291] and our results may not be generalizable in other settings. The use of a GRCS for the assessment of patient status has been criticized as being biased.[147] However, few if any better alternatives have been determined. Finally, all predictive analyses were short-term. The predictive properties of the test items in this study may be different for a longer period of time.

This study supports the contentions[31] and findings[28.292] of others that information contained the patient history is often the most powerful diagnostic tool available and often supplies the clinician with everything needed to clinch a diagnosis. The clinical examination test items with the most powerful LR value for subjects who underwent surgery and worsened were questions. Often, multiple questions had highly predictive LR values. Indeed, based on the reliability and diagnostic accuracy coefficients, the

"hardest" data in this study were those usually thought to be "soft".[289] This study also supports the observation that when multiple, appropriate test items are combined into a single TIC, diagnostic power is often greatly improved.[277p 132]

The vast majority of clinical examination items in this study was found to have acceptable reliability and is most likely due to operationally defining our tests and standardizing the examination protocol. Two clinical examination items that are of particular interest are question 7-"Sympotms keep from sleep.." and the ULTT A for the diagnosis of CR. Both LR values were acceptable for Question 7 and its LR+= 6.5 was the only single test item value in this study that was definitive. The ULTT was perfectly sensitive (95CI= .87 – 1.0) and had an LR-= .15. By comparison, the computed LR- value of the straight-leg raise, a commonly used screening test for lumbar radiculopathy, is only .33. Further investigation may demonstrate that this CR TIC, a simple question, or a single examination procedure can be very powerful and inexpensive tools useful for the diagnosis and screening of patients with CR.

The majority of individual test items and TIC for the CTS subjects were less powerful than were the tests and TIC for the diagnosis of CR. Spectrum bias did appear to have an impact of the diagnostic characteristics of the tests for the CTS group as a whole. Interestingly, the two test items with the best diagnostic power for CTS were the HSAM's. These two scales were developed for evaluative purposes, which usually diminishes the discriminative properties of the scales.[142] This was not the case in our study.

Unfortunately, the predictive validity of clinical examination items for CR were unable to be assessed in this study. The predictive validity of clinical examination items for CTS will need to be tested in different settings and over a longer period of time. Longer follow-up times are required to determine which subjects with CTS will improve as the natural course of this condition may prolonged and in many cases unfavorable.[15]

A replication study, in part or whole, with a larger sample is required to improve the precision of point estimates obtained for the clinical examination test items in this study. The TIC's developed in this study must be examined in an independent sample before their utility can be adequately assessed. *Clinimetrics* involves the quantifying of data that are observed, judged, and decided on during the clinical examination by clinicians themselves[290pp 1 - 5] This work should be regarded as an attempt toward developing *Clinimetrics* for two common musculoskeletal disorders regularly encountered in clinical practice. The tests items and TIC's developed and assessed in this study are sensible, have face validity, include a formal expression of the index of interest, and are certainly easy to use.[290pp 1 - 5. 141 - 166] Further work is needed to establish and validate the content of the clinical examination and TIC's used in this study.

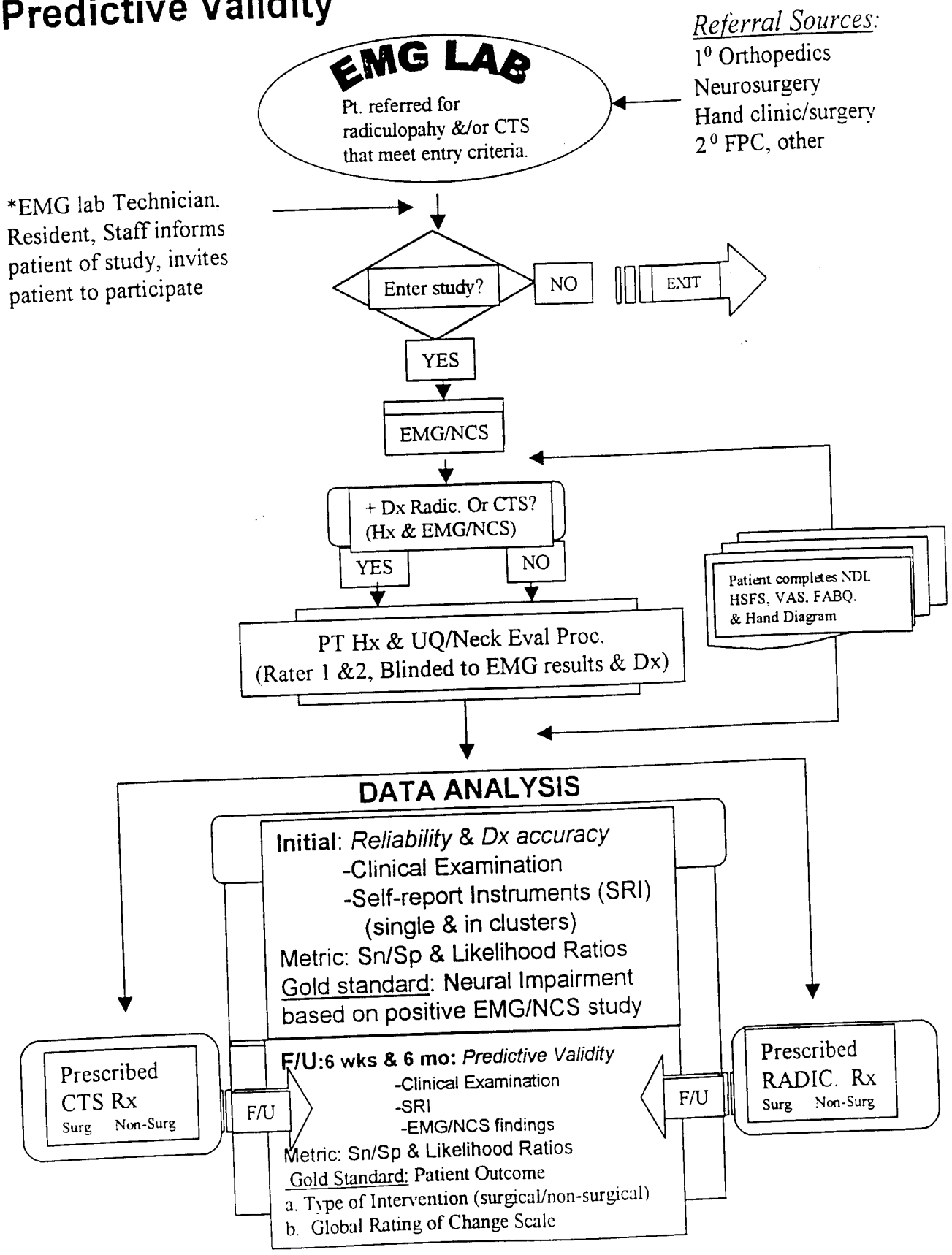
7.6 Conclusion

Of the 54 clinical examination tests in this study, the majority were found to have Fair to Good or Good reliability and a few had Excellent reliability; the reliability of 11 of these items was poor to moderate. Two CR test items had a definitively acceptable level of Sn while no CTS test items were definitively acceptable. A number of CR clinical examination items had a definitively acceptable level of Sp and three CTS test items were found to have a definitively acceptable level of Sp. Likelihood ratio point estimates indicated that several test items and TIC's are potentially useful for the diagnosis and prognosis of CR and CTS. Question number 7 was found to have a definitively acceptable LR+, as were the CR TIC and the surgery TIC. None of the definitive LR+ values had lower bounds that would result in post-test probabilitie changes larger than 33%.

APPENDICES

Appendix A

Cervical Radiculopathy and CTS: A Prospective Study of Test Reliability, Diagnostic Accuracy and Predictive Validity



Multi-center Study of Cervical Radiculopathy and Carpal Tunnel Syndrome (CTS)

Will you consider being a participant in our study? The physical therapy clinic and EMG lab are trying to determine the usefulness of commonly used clinical examination procedures and questionnaires that are used to diagnose patients with your suspected condition (either cervical radiculopathy and/or CTS). To compensate you for your time, you will be paid up to $25.00 for returning data collection forms.

Specifically, we want to determine the reliability and accuracy of the procedures and questionnaires and find out whether they can help us predict how well you will respond to treatment. Often times more expensive and invasive tests (such as EMG) can be avoided if clinical tests for detecting the condition have been shown to be reliable and accurate.

If you choose to help us, you will fill out some questionnaires and a standardized EMG examination of your affected limb will be performed. Having a "standardized" EMG examination means two extra nerve conduction procedures will be performed and possibly 1-4 extra muscles will be examined with an EMG pin electrode. We say "possibly" because the muscles to be tested during the standardized examination are often tested anyway when performing an EMG examination on patients with your suspected condition.

Following your EMG examination, a physical therapist will ask you some questions about your condition, take some measurements of your neck and wrists, and perform several tests designed to alter your symptoms. After a brief rest period, a second physical therapist will repeat the same examination. Finally, at six weeks and six months from now you will be mailed four of the same questionnaires you filled out today and a scale asking you rate your improvement along with a self-addressed stamped return envelope. Once the we receive the forms, paperwork will be processed to issue you a check in your name which you will receive in the mail: $15.00 for the 6 week forms and $10.00 for the 6 month forms. You will incur absolutely no additional cost if you choose to participate in this study.

Well, in a nut-shell that's what your participation in this study will involve. The technicians, therapists, or doctors involved in your care are happy to answer any questions you may have. And, of course, a more detailed description of the study is contained in the patient consent form that you must read and sign before you participate.

Thank you very much for considering participating in this important study. Your assistance will help us answer our questions. The answers we obtain could help future sufferers of cervical radiculopathy and CTS by providing a way for their condition to be detected more quickly and comfortably and predicting how they may respond to certain kinds of treatment.

Screening Questions for Study Participation:

Name_____ Age___

1. Referred for: Cervical radiculopathy ☐ CTS ☐ Side affected: L☐ R☐, Dom☐ NonDom☐
2. When did this current episode of symptoms first begin?_____
3. Is this the first time you have experienced the symptoms of this conditon? Yes ☐ No ☐
4. If no, how many previous episodes of this condition have you experienced in the past?

   Only one other time ☐ Two to Three times ☐ Four or more times ☐
5. **Have you ever had an EMG examination of your affected limb in the past for this same condition ?** **Yes**☐ No ☐
6. Have you received treatment or are you currently being treated for this condition? Yes☐ No ☐

   (check all that apply): cervical traction ☐ exercise ☐ wrist splint ☐ injections ☐ other ☐
7. Do you take medications for your upper limb symptoms or any other condition? Yes☐ No☐

   (If yes, list medications:_____)

   (_____)
8. Have you ever had 6 or more drinks during the course of your workday?

Carefully read the following list and check the box that applies to you:

Do you have any of the following conditions?

	Yes	No
Diabetes	☐	☐
Thyroid disease	☐	☐
Cancer	☐	☐
Kidney disease	☐	☐
Rheumatoid arthritis	☐	☐
Any history of stroke or neurological disease	☐	☐
Blood clotting disorder	☐	☐
Exposure to lead, mercury, or industrial solvents	☐	☐
Prior fractures of your affected hand, wrist or your neck	☐	☐
Prior surgery of your affected hand, wrist, or your Neck	☐	☐
Numbness and tingling of both arms and/or both legs	☐	☐
Pregnant	☐	☐
Off work for longer than 6 months due to your symptoms	☐	☐

Appendix B

63491

Multicenter Cervical Radiculopathy & CTS Study

D/M/Y ☐☐ / ☐☐ / 19☐☐

Subject ID # ☐☐ Facility # ☐ Form # 1

196

Facility number
1=Pitt
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

Last Name

First Name

Address

Address

City

Home Phone ☐☐☐ - ☐☐☐ - ☐☐☐☐

Email

Age ☐☐

Gender ○ Male ○ Female

Is there Workers compensation involved in your case? ○ No ○ Yes

Is there litigation pending or a settlement involved in your case? ○ No ○ Yes

State ☐☐ Zip ☐☐☐☐☐

Work Phone ☐☐☐ - ☐☐☐ - ☐☐☐☐

Contact Person or Relative Not LIving with You

Last Name

First Name

Address

Address

City

Home Phone ☐☐☐ - ☐☐☐ - ☐☐☐☐

Email

State ☐☐ Zip ☐☐☐☐☐

Work Phone ☐☐☐ - ☐☐☐ - ☐☐☐☐

5532

NDI FORM

Eval Period: ○ Initial ○ 6 weeks ○ 6 months

D/M/Y [ ][ ] / [ ][ ] / 1 9 [ ][ ]

Subject ID # [ ][ ] Facility # [ ] Form # 8

1=Pitt
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

Please Read: This Questionnaire is designed to give the doctor information as to how your neck pain has affected your ability to manage in Please answer every section and mark in each section only the ONE box which applies to you. We realize you may consider that two of tl any one section may relate to you, but please just mark the box which most closely describes your problem.

Section 1 - Pain Intensity

- ○ I have no pain at the moment.
- ○ The pain is very mild at the moment.
- ○ The pain is moderate at the moment.
- ○ The pain is fairly severe at the moment.
- ○ The pain is very severe at the moment.
- ○ The pain is the worst imaginable at the moment.

Section 2 - Personal Care

- ○ I can look after myself normally without causing extra
- ○ I can look after myself normally but it causes extra pain.
- ○ It is painful to look after myself and I am slow and careful.
- ○ I need some help but manage most of my personal care.
- ○ I need help everyday in most aspects of self care.
- ○ I do not get dressed, I wash with difficulty and stay in bed.

Section 3 - Lifting

(Skip if you have not attempted liftin since the onset of your neck pain)

- ○ I can lift heavy weights without extra pain
- ○ I can lift heavy weights but it causes extra pain
- ○ Pain prevents me lifting heavy wieghts off the floor, but I can manage if they are conveniently positioned, e.g. on a table
- ○ Pain prevents me lifting heavy wieghts but I can manage light to medium wieghts if they are conveniently positioned.
- ○ I can only lift very light weights.
- ○ I cannot lift or carry anything at all.

Section 4 - Reading

- ○ I can read as much as I want to with no pain in my neck
- ○ I can read as much as I want to with slight pain in my
- ○ I can read as much as I want with moderate pain in my neck.
- ○ I can't read as much as I want because of moderate pain in my neck.
- ○ I can hardly read at all because of severe pain in my neck
- ○ I cannot read at all.

Section 5 - Headaches

- ○ I have no headaches at all.
- ○ I have slight headaches which come infrequently.
- ○ I have moderate headaches which come infrequently.
- ○ I have moderate headaches which come frequently.
- ○ I have severe headaches which come frequently.
- ○ I have headaches all the time.

Section 6 - Concentration

- ○ I can concentrate fully when I want to with no difficulty.
- ○ I can concentrate fully when I want to with slight difficulty.
- ○ I have a fair degree of difficulty in concentrating when I want to.
- ○ I have a lot of difficulty in concentrating when I want to.
- ○ I have a great deal of difficulty in concentrating when I want to.
- ○ I cannot concentrate at all.

Section 7 - Work

- ○ I can do as much work as I want to.
- ○ I can only do my usual work, but no more.
- ○ I can do most of my usual work, but no more.
- ○ I cannot do my usual work.
- ○ I can hardly do any work at all.
- ○ I can't do any work at all.

Section 8 - Driving

- ○ I can drive my car without any neck pain.
- ○ I can drive my car as long as I want with slight pain in neck.
- ○ I can drive my car as long as I want with moderate pai my neck.
- ○ I can't drive my car as long as I want because of mode pain in my neck.
- ○ I can hardly drive at all because of severe pain in my r
- ○ I can't drive my car at all.

Section 9 - Sleeping

- ○ I have no trouble sleeping
- ○ My sleep is slightly disturbed(less than 1 hr. sleepless)
- ○ My sleep is mildly disturbed (1-2 hrs. sleepless)
- ○ My sleep is moderayely disturbed(2-3 hrs. sleepless)
- ○ My sleep is greatly disturbed(3-5 hrs. sleepless)
- ○ My sleep is completely disturbed(5-7 hrs. sleepless)

Section 10 - Recreation

- ○ I am able to engage in all my recreation activities with no neck pain at all.
- ○ I am able to engage in all my recreation activities, with some pain in my neck.
- ○ I am able to engage in most, but not all of my usual recreation activities because of pain in my neck.
- ○ I am able to engage in a few of my usual recreation activities because of pain in my neck.
- ○ I can hardly do any recreation activities because of pain in my neck
- ○ I can't do any recreation activities at all.

Note: Please give only 1 Answer per question

5340

Eval Period: ○ Initial ○ 6 weeks ○ 6 months

D/M/Y ☐☐ / ☐☐ / 19☐☐

Subject ID # ☐☐ **Facility #** ☐ **Form #** 10

Facility number
1=Pitt 198
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

SX Severity Scale

The following questions refer to your symptoms for a typical 24 hr. period during the last two weeks (circle one answer to each question).

1. How severe is the hand or wrist pain that you have at night?
○ I do not have hand or wrist pain at night
○ Mild pain
○ Moderate pain
○ Severe pain
○ Very severe pain

2. How often did hand or wrist pain wake you up during a typical night in the past two weeks?
○ Never
○ Once
○ Two to three times
○ Four or five times
○ More than five times

3. Do you typically have pain in your hand or wrist during the daytime?
○ I never have pain during the day.
○ I have mild pain during the day.
○ I have moderate pain during the day.
○ I have severe pain during the day.
○ I have very severe pain during the day.

4. How often do you have hand or wrist pain during the daytime?
○ Never
○ Once or twice a day.
○ Three to five times a day.
○ More than five times a day.
○ The pain is constant.

5. How long, on average, does an episode of pain last during the daytime?
○ I never get pain during the day.
○ Less than 10 minutes.
○ 10 to 60 minutes.
○ Greater than 60 minutes.
○ The pain is constant throughout the day.

6. Do you have numbness (loss of sensation) in your hand?
○ No
○ I have mild numbness
○ I have moderate numbness.
○ I have severe numbness.
○ I have very severe numbness.

7. Do you have weakness in your hand or wrist ?
○ No weakness
○ Mild weakness.
○ Moderate weakness.
○ Severe weakness.
○ Very severe weakness.

8. Do you have tingling sensations in your hand?
○ No tingling.
○ Mild tingling.
○ Moderate tingling
○ Severe tingling.
○ Very severe tingling.

9. How severe is numbness (loss of sensation) or tingling at night?
○ I have no numbness or tingling at night.
○ Mild
○ Moderate
○ Severe
○ Very Severe

10. How often did hand numbness or tingling wake you up during a typical night during the past two weeks?
○ Never
○ Once
○ Two or three times
○ Four or five times
○ More than five times

11. Do you have difficulty with the grasping and use of small objects such as keys or pens?
○ No difficulty
○ Mild difficulty
○ Moderate difficulty
○ Severe difficulty
○ Very severe difficulty

Functional Status Scale

199

Facility number
1=Pitt
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

Eval Period: ○ Initial ○ 6 weeks ○ 6 months

D/M/Y ☐☐ / ☐☐ / 19☐☐

Subject ID # ☐☐ **Facility #** ☐ **Form #** 11

On a typical day during the past two weeks have hand and wrist symptoms caused you to have any difficulty doing the activities listed below? Please fill-in the bubble beside the number that best describes your ability to do the activity

Activity	No Difficulty	Mild Difficulty	Moderate Difficulty	Severe Difficulty	Cannot do at all due to hand or wrist symptoms
Writing	○ 1	○ 2	○ 3	○ 4	○ 5
Buttoning of clothes	○ 1	○ 2	○ 3	○ 4	○ 5
Holding a book while reading	○ 1	○ 2	○ 3	○ 4	○ 5
Gripping of a telephone handle	○ 1	○ 2	○ 3	○ 4	○ 5
Opening of jars	○ 1	○ 2	○ 3	○ 4	○ 5
Household chores	○ 1	○ 2	○ 3	○ 4	○ 5
Carrying of grocery bags	○ 1	○ 2	○ 3	○ 4	○ 5
Bathing and dressing	○ 1	○ 2	○ 3	○ 4	○ 5

5616

5806

HAND DIAGRAM & VAS FORM

Eval Period: O Initial O 6 weeks O 6 months

D/M/Y ☐☐ / ☐☐ / 19☐☐

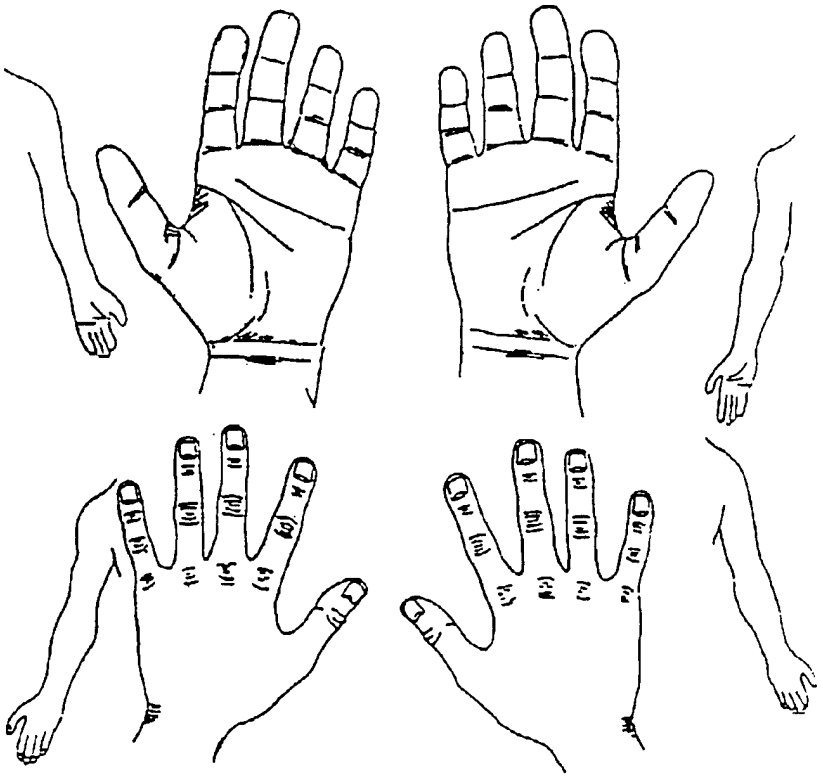
Subject ID # ☐☐ **Facility #** ☐ **Form #** 9

Facility numbe
1=Pitt200
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

Using the key shown below, please draw in on the diagram the areas of each hand/arm where you have felt pain, numbness, tingling, or other types of discomfort on a **typical day during the PAST TWO WEEKS.**

Key:
XXXX .....Pain
//// ....Numbness/tingling
0000 ....Other discomfort (please describe: ☐

LEFT HAND/ARM **RIGHT HAND/ARM**



Please use the three scales below to rate your pain over the past 24 hours. Use the upper line to rate your pain level **right now.**

Use the other scales to rate your pain at its **worst** and **best** over the past 24 hours.

Office use (in mm)
☐☐☐
☐☐☐
☐☐☐

Right Now:
NO PAIN _____ **WORST POSSIBLE PAIN**

Worst past 24 hrs.
NO PAIN _____ **WORST POSSIBLE PAIN**

Best past 24 hrs.
NO PAIN _____ **WORST POSSIBLE PAIN**

5411

FABQ (ACTIVITY)

Eval Period: ○ Initial ○ 6 weeks ○ 6 months

D/M/Y [ ][ ] / [ ][ ] / [1][9][ ][ ]

Subject ID # [ ][ ]

Facility # [ ]

Form # [1][2]

Facility numbe
1=Pitt
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

201

Here are some of the things other patients have told us about their pain. For each statement please mark the number from 0 to 6 to indicate how much physical activities such as bending, lifting, walking or driving affect or would affect your neck and/or hand pain.

	Completely Disagree			Unsure			Completely Agree
1. My pain was caused by physical activity	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
2. Physical activity makes my pain worse	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
3. Physical activity might harm my neck an wrist	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
4. I should not do physical activities whicl (might) make my pain worse	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
5. I cannot do physical activities which (might) make my pain worse	○ 0	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6

5476

FABQ (WORK)

Eval Period: ○ Initial ○ 6 weeks ○ 6 months

Date ☐☐ / ☐☐ / 19☐☐

Subject ID # ☐☐ **Facility #** ☐ **Form #** 13

Facility number 202
1= Pitt
2= NNMC
3= WHMC
4= AFA

The following statements are about how your normal work affects or would affect your neck and/or hand pain.

	Completely Disagree			Unsure			Completely Agree
6. My pain was caused by my work or by accident at work	○0	○1	○2	○3	○4	○5	○6
7. My work aggravated my pain	○0	○1	○2	○3	○4	○5	○6
8. I have a claim for compensation for my pain	○0	○1	○2	○3	○4	○5	○6
9. My work is too heavy for me	○0	○1	○2	○3	○4	○5	○6
10. My work makes or would make my pain worse	○0	○1	○2	○3	○4	○5	○6
11. My work might harm my neck and/or hand	○0	○1	○2	○3	○4	○5	○6
12. I should not do my regular work with my present pain	○0	○1	○2	○3	○4	○5	○6
13. I cannot do my normal work with my present pain	○0	○1	○2	○3	○4	○5	○6
14. I cannot do my normal work until my pain is treated	○0	○1	○2	○3	○4	○5	○6
15. I do not think that I will be back to my normal work within 3 months	○0	○1	○2	○3	○4	○5	○6
16. I do not think that I will ever be able to go back to that work	○0	○1	○2	○3	○4	○5	○6

53208

Eval Period: D/M/Y ☐☐ / ☐☐ / ☐☐☐☐

○ 6 weeks

○ 6 months

Subject ID # ☐☐ Facility # ☐ Form # 14

203

Facility number
1=Pitt
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

FOLLOW-UP & OUTCOME

1. Since your EMG/NCS test, have you **HAD** surgery? ○ No ○ Yes

2. Since your EMG/NCS test, have you BEEN OFFERED surgery? ○ No ○ Yes

3. Compared to your condition prior to treatment, which item on the scale below **best** describes your present condition (choose only one):

Patient Global Rating Scale

○ A very great deal worse
○ A great deal worse
○ Quite a bit worse
○ Moderately worse
○ Somewhat worse
○ A little bit worse
○ A tiny bit worse (almost the same)
○ About the same
○ A tiny bit better (almost the same)
○ A little bit better
○ Somewhat better
○ Moderately better
○ Quite a bit better
○ A great deal better
○ A very great deal better

4. Please check any of the following treatments you have received for your condition (check all that apply):

Medication

○ None ○ Anti-inflammitory (Motrin, Advil, Naproxsyn, etc.) ○ Narcotics ○ Tylenol ○ Steroids ○ Other

Other Conservative Treatments:

○ None ○ Collar ○ Traction ○ Manipulation ○ Exercise ○ Wrist Splint ○ Injection

5. Please check any of the following treatments you are **still receiving** for your condition (check all that apply):

Medication

○ None ○ Anti-inflammitory (Motrin, Advil, Naproxsyn, etc.) ○ Narcotics ○ Tylenol ○ Steroids ○ Other

Other Conservative Treatments:

○ None ○ Collar ○ Traction ○ Manipulation ○ Exercise ○ Wrist Splint ○ Injection

Is there anything else you would like to tells us about your condition?

Appendix C

5115

QUESTION FORM

D/M/Y ☐☐ / ☐☐ / 19☐☐

Subject ID # ☐☐ Facility # ☐ Rater# ☐ Exam# 1 Form # 6

Facility number
1=Pitt
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

205

1. **Which** of the following symptoms are most bothersome for you? (choose one)
   - O Pain
   - O Numbness & Tingling
   - O Loss of feeling
2. **Where** are your symptoms most bothersome?
   - O Neck
   - O Shoulder or shoulder blade
   - O Arm above elbow
   - O Arm below elbow
   - O Hands and/or fingers
3. Which of the following best describes the behaviour of your symptoms?
   - O Constant
   - O Intermittent (symptoms come & go)
   - O Variable (symptoms improve or worsen at times)

4. Does your affected hand feel "fat" or "swollen"? O NO O YES
5. Do you have trouble with fumbling or dropping objects from your affected hand? O NO O YES
6. Does your entire affected limb and/or hand feel numb? O NO O YES
7. Do your symptoms **keep** you from falling asleep at night? O NO O YES
8. Do your symptoms **wake** you during the night? O NO O YES
9. Do your symptoms improve with moving or positioning your **neck**? O NO O YES
10. Do your symptoms improve with moving, "shaking", or positioning your **wrist or hands**? O NO O YES
11. Are your symptoms brought on or made worse when performing tasks that require a lot or grasping or hand and/or finger use? O NO O YES

5140

QUESTION FORM

D/M/Y ☐☐ / ☐☐ / 19☐☐

Subject ID # ☐☐ Facility # ☐ Rater# ☐ Exam# 2 Form # 6

Facility number
1=Pitt 206
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

1. **Which** of the following symptoms are most bothersome for you? (choose one)
   - O Pain
   - O Numbness & Tingling
   - O Loss of feeling

2. **Where** are your symptoms most bothersome?
   - O Neck
   - O Shoulder or shoulder blade
   - O Arm above elbow
   - O Arm below elbow
   - O Hands and/or fingers

3. Which of the following best describes the behaviour of your symptoms?
   - O Constant
   - O Intermittent (symptoms come & go)
   - O Variable (symptoms improve or worsen at times)

4. Does your affected hand feel "fat" or "swollen"? O NO O YES

5. Do you have trouble with fumbling or dropping objects from your affected hand? O NO O YES

6. Does your entire affected limb and/or hand feel numb? O NO O YES

7. Do your symptoms **keep** you from falling asleep at night? O NO O YES

8. Do your symptoms **wake** you during the night? O NO O YES

9. Do your symptoms improve with moving or positioning your **neck**? O NO O YES

10. Do your symptoms improve with moving, "shaking", or positioning your **wrist or hands**? O NO O YES

11. Are your symptoms brought on or made worse when performing tasks that require a lot or grasping or hand and/or finger use? O NO O YES

4012

EXAM FORM

D/M/Y ☐☐ / ☐☐ / 19☐☐

Subject ID # ☐☐ Facility # ☐ Rater# ☐ Exam# 1 Form # 7

Facility number 207
1=Pitt
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

Involved Side

○ Left ○ Right

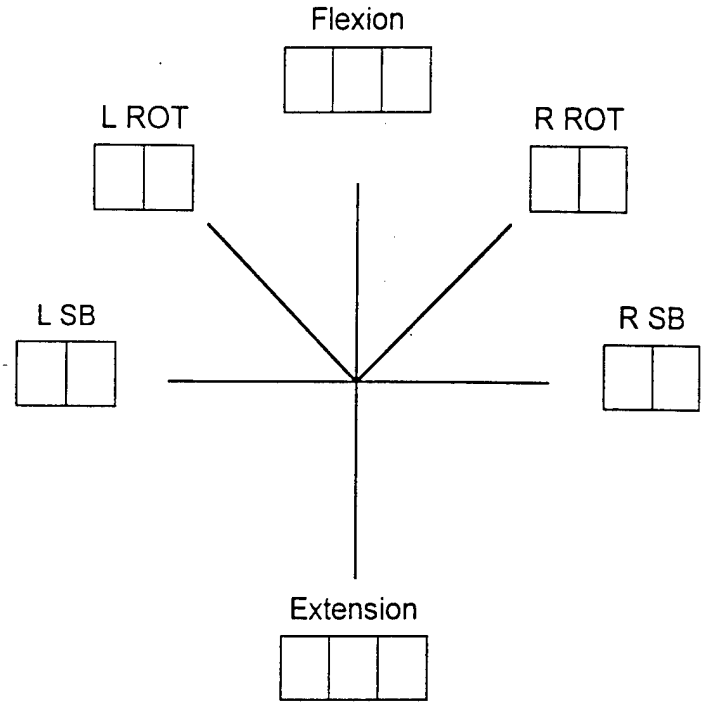
Measurements:

Wrist Ratio:

Ant-Post (mm): ☐☐☐

Med-Lat (mm): ☐☐☐

Cervical ROM



Provocative Tests:

	NEG	POS	NOT TOL.
Spurling's A	○	○	○
Spurling's B	○	○	○
Shoulder Abduction	○	○	○
Valsalva	○	○	○
Distraction	○	○	○
Tinel's A	○	○	○
Tinel's B	○	○	○

ULTT GRADING: 0=Negative;
A: 1-6,s/s diff.=Positive B: 1-5,s/s diff.=Positive

								s/s diff.	NOT TOL.
ULTT A	○0	○1	○2	○3	○4	○5	○6	○	○
ULTT B	○0	○1	○2	○3	○4	○5		○	○

Timed Tests:

CCT (30 sec.)
- ○ **Negative**
- ○ **Pos. <15**
- ○ **Pos. <30**
- ○ **Not Tol.**

Phalen's (60 sec.)
- ○ **Negative**
- ○ **Pos. <15**
- ○ **Pos. <30**
- ○ **Pos. <45**
- ○ **Pos. <60**
- ○ **Not Tol.**

NOTES:

4108

EXAM FORM

D/M/Y ☐☐ / ☐☐ / 19☐☐

Subject ID # ☐☐ Facility # ☐ Rater# ☐ Exam# 2 Form # 7

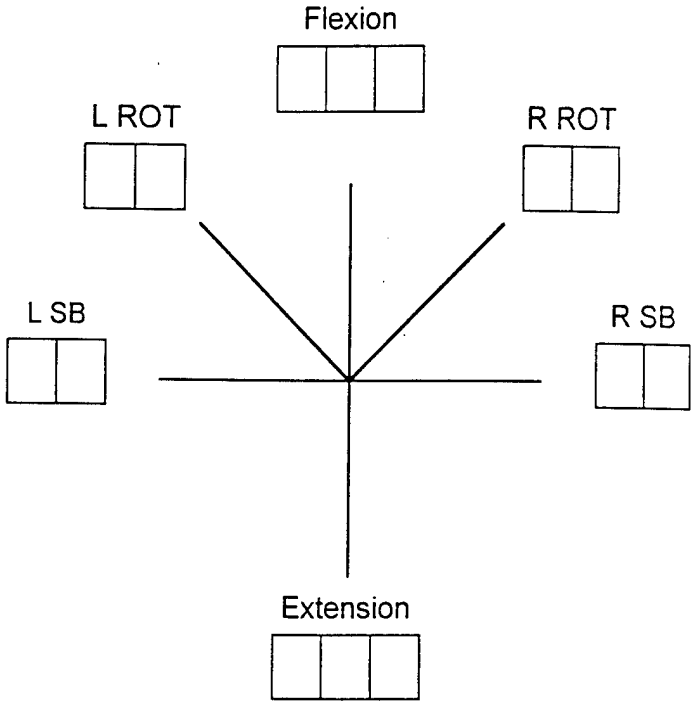
Facility number 208
1=Pitt
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

Involved Side

○ Left ○ Right

Measurements:

Cervical ROM



Wrist Ratio:

Ant-Post (mm): ☐☐☐

Med-Lat (mm): ☐☐☐

Provocative Tests:

	NEG	POS	NOT TOL.
Spurling's A	○	○	○
Spurling's B	○	○	○
Shoulder Abduction	○	○	○
Valsalva	○	○	○
Distraction	○	○	○
Tinel's A	○	○	○
Tinel's B	○	○	○

ULTT GRADING: 0=Negative;
A: 1-6,s/s diff.=Positive B: 1-5,s/s diff.=Positive

								s/s diff.	NOT TOL.
ULTT A	○0	○1	○2	○3	○4	○5	○6	○	○
ULTT B	○0	○1	○2	○3	○4	○5		○	○

Timed Tests:

CCT (30 sec.)
- ○ **Negative**
- ○ **Pos. <15**
- ○ **Pos. <30**
- ○ **Not Tol.**

Phalen's (60 sec.)
- ○ **Negative**
- ○ **Pos. <15**
- ○ **Pos. <30**
- ○ **Pos. <45**
- ○ **Pos. <60**
- ○ **Not Tol.**

NOTES:

4337

Neurologic Exam

D/M/Y ☐☐ / ☐☐ / 19☐☐

Subject ID # ☐☐ | Facility # ☐ | Rater# ☐ | Exam# 1 | Form # 5

Facility number
1=Pitt
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

209

1. MSRs

	Left				Right			
	Absent	Reduced	WNL	Increased	Absent	Reduced	WNL	Increased
Biceps	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
Brachioradialis	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
Triceps	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3

2. Sensory Examination
(paperclip point)

	Left				Right			
Dermatomes	Absent	Reduced	WNL	Increased	Absent	Reduced	WNL	Increased
C5 (lateral deltoid)	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C6 (rad. aspect of index f.)	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C7 (dorsum middle f.)	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C8 (med. aspect little f.)	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
T1 (med. aspect mid-forearm)	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
Median N. Distribution (Palmar surface compared to thenar skin)	**Left**				**Right**			
Thumb	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
Index Finger	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
Middle Finger	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3

3. Motor Examination:

Myotome	Muscle	Absent	Markedly Reduced (P- to F)	Reduced (F+ to G)	Normal (N)	Absent	Markedly Reduced (P- to F)	Reduced (F+ to G)	Normal (N)
C5	deltoid	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C6	biceps	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C7	ext carp rad	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
	triceps	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
	flex carp rad	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C8	abd poll brev	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
T1	first dorsal Int	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3

4979

Neurologic Exam

D/M/Y ☐☐ / ☐☐ / 19☐☐

Subject ID # ☐☐ Facility # ☐ Rater# ☐ Exam# 2 Form # 5

Facility number
1=Pitt
2=BAMC
3=WHMC
4=AFA
5=MGMC
6=FT Camp.

210

1. MSRs

	Left				Right			
	Absent	Reduced	WNL	Increased	Absent	Reduced	WNL	Increased
Biceps	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
Brachioradialis	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
Triceps	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3

2. Sensory Examination:
(paperclip point)

	Left				Right			
Dermatomes	Absent	Reduced	WNL	Increased	Absent	Reduced	WNL	Increased
C5 (lateral deltoid)	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C6 (rad. aspect of index f.)	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C7 (dorsum middle f.)	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C8 (med. aspect little f.)	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
T1 (med. aspect mid-forearm)	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
Median N. Distribution (Palmar surface compared to thenar skin)	**Left**				**Right**			
Thumb	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
Index Finger	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
Middle Finger	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3

3. Motor Examination:

Myotome	Muscle	Absent	Markedly Reduced (P- to F)	Reduced (F+ to G)	Normal (N)	Absent	Markedly Reduced (P- to F)	Reduced (F+ to G)	Normal (N)
C5	deltoid	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C6	biceps	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C7	ext carp rad	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
	triceps	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
	flex carp rad	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
C8	abd poll brev	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3
T1	first dorsal Int	O 0	O 1	O 2	O 3	O 0	O 1	O 2	O 3

Bibliography

1. Kelsey JL. Epidemiology of Musculoskeletal Disorders. NY: Oxford University Press, Inc., 1982.

2. Jette AM, Delitto A. Physical therapy treatment choices for musculoskeletal impairments. Phys Ther. 1997; 77:145-154.

3. Aryanpur J, Ducker TB. Differential Diagnosis and Management of Cervical Spine Pain. In: Tollison CD. editor. Handbook of Chronic Pain Management. Baltimore: Williams & Wilkins, 1989:

4. Wilson PR. Chronic neck pain and cervicogenic headache. Clin J Pain 1991; 7:5-11.

5. Gore DR, Sepic SB, Gardner GM, Murray MP. Neck pain: a long-term follow-up of 205 patients. Spine 1987; 12:1-5.

6. Ellenberg MR, Honet JC, Treanor WJ. Cervical Radiculopathy. Arch Phys Med Rehabil 1994; 75:342-352.

7. Connell MD, Wiesel SW. Natural history and pathogenesis of cervical disc disease. Orthop Clin North Am 1992; 23:369-380.

8. Heller JG. The syndromes of degenerative cervical disease. Orthop Clin North Am 1992; 23:381-394.

9. Erdil M, Maurer KF, Dickerson OB. The Burden of Cumulative Trauma Disorders. In: Erdil M, Dickerson OB, editors. Cumulative Trauma Disorders. New York: Van Nostrand Reinhold, 1997:

10. Sevens JC, Sun S, Beard CM, O'Fallon WM, Kurland LT. Carpal tunnel syndrome in Rochester, Minnesota. 1961 to 1980. Neurology 1988; 38:134

11. Franklin GM, Haug J, Heyer N, Checkoway H, Peck N. Occupational carpal tunnel syndrome in Washington State. Am J of Public Health 1991; 81:741

12. Masear VR, Hayes JM, Hyde AG. An industrial cause of carpal tunnel syndrome. J Hand Surg [Am] 1986; 11:222-227.

13. Katz JN, Gelberman RH, Wright EA, Abrahamson SO, Lew RA. A preliminary scoring system for assessing the outcome of carpal tunnel relese. J Hand Surg [Am] 1994; 19:531

14. Mackinnon SE. Secondary carpal tunnel surgery. Neurosurg Clin N Am 1991; 2:75-91.

15. Weiss AP, Sachar K, Gendreau M. Conservative management of carpal tunnel syndrome: a re-examination of steroid injection and splinting. J Hand Surg [Am] 1994; 19A:410-415.

16. Katz JN, Larson MG, Sabra A, Krarup C, Stirrat CR, Sethi R, et al. The carpal tunnel syndrome: diagnostic utility of the history and physical examination findings. Ann Int Med 1990; 112:321-327.

17. Yoss RE, Corbin KB, MacCarty CS, Love JG. Significance of symptoms and signs in localization of involved root in cervical disc protrusions. Neurology 1999; 7:673-683.

18. Hurst LC, Weissberg D, Carroll RE. The relationship of double crush to carpal tunnel syndrome (an analysis of 1,000 cases of carpal tunnel syndrome). J Hand Surg [Am] 1985; 10:202-204.

19. Pfeffer G, Osterman AL. Double crush syndrome: cervical radiculopathy and carpal tunnel syndrome. J Hand Surg [Am] 1986; 11:766-766.

20. Upton AR, McComas AJ. The double crush in nerve entrapment syndromes. Lancet 1973; 2:359-362.

21. Viikari-Juntura E, Porras M, Laasonen EM. Validity of clinical tests in the diagnosis of root compression in cervical disc disease. Spine 1989; 14:253-257.

22. MacDermid J. Accuracy of clinical tests used in the detection of carpal tunnel syndrome: a literature review. J Hand Ther 1991; 169-176.

23. Holleman DR, Simel DL. Quantitative assessments from the clinical examination: how should clinicians integrate the numerous results? J Intern Med 1997; 12:165-171.

24. Sackett DL, Haynes RB, Guyett GH. Clinical Epidemiology: A Basic Science for Clinical Medicine. 2nd Ed. ed. Boston, MA: Little, Brown, and Company, Inc., 1991.

25. Marx RG, Bombardier C, Wright JG. What do we know about the reliability and validity of physical examination tests used to examine the upper extremity? J Hand Surg [Am] 1999; 24A:185-193.

26. Wehrle J. Chronic wrist injuries associated with repetitive hand motions in industry. 1976; The University of Michigan, Ann Arbor, Michigan: Department of Industrial and Operations Engineering. Occupational Health and Safety Technical Report.

27. Simel DL, Dujardin B. The clinical examination: an agenda to make it more rational. JAMA 1997; 277:572-574.

28. Hampton JR, Harrison MJG, Mitchell JRA, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and managment of medical outpatients. BMJ 1975; 2:486-489.

29. Sandler G. The importance of the history in the medical clinic and the cost of unnecessary tests. Am Heart J 1980; 100:928-931.

30. Crombie DL. Diagnostic Process. J Coll Gen Pract 1963; 6:579-589.

31. Sackett DL, Drummond R. The science of the art of the clinical examination. JAMA 1992; 267:2650-2652.

32. Bush B, Shaw S, Cleary P. Screening for alcohol abuse using the CAGE questionnaire. Am J Med 1987; 82:231

33. Sharon E.Straus, Finlay McAlister David L. Sackett Director. Clinical Assessment of the Reliability of the Clinical Examination. NHS R&D Centre for Evidence-Based Medicine, Nuffield Department of Medicine, University of Oxford, Radcliffe Hospital, Oxford, England, 1999; (Internet Communication).

34. Magee DJ. Orthopedic Physical Assessment. Philadelphia, PA: WB Saunders Company, 1992.

35. Richardson JK, Iglarsh ZA. Clinical Orthopedic Physical Therapy. Philadelphia, PA: W.B. Saunders Company, 1994.

36. Stratford P, Spadoni G, Berk A. Selecting the best clinical diagnostic test. Orthopaedic Practice 1999; 11:31-35.

37. Schwartz J. Evaluating diagnostic tests: what is done---what needs to be done. J Gen Int Med 1986; 1:266-267.

38. Sackett DL, Haynes RB, Guyett GH. Clinical Epidemiology: A Basic Science for Clinical Medicine. 2nd Ed. ed. Boston, MA: Little, Brown, and Company, Inc., 1991.

39. Fineberg HV, Bauman R, Sosman M. Computerized cranial tomography: effect on diagnostic and therapeutic plans. JAMA 1977; 238:224-230.

40. Deyo RA, Haselkorn J, Hoffman R, Kent DL. Designing studies of diagnostic tests for low back pain or radiculopathy. Spine 1994; 19:2057S-2065S.

41. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. Norwalk, CT: Appleton & Lange, 1993.

42. Sackett DL, Wennberg JE. Choosing the best research design for each question: It's time to stop squabbling over the "best" methods. BMJ 1997; 315:1636

43. Sheldon TA, Guyatt G, Haines A. Getting research findings into practice: when to act on the evidence. BMJ 1998; 317:139-142.

44. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Eng J Med 1978; 299:926-930.

45. Jaeschke R, Guyatt G, Sackett DL. Users' guide to the medical literature. III. How to use an article about a diagnostic test: A. Are the results of the study valid? JAMA 1994; 271:389-391.

46. Mulrow C, Linn WD, Gaul MK, Pugh JA. Assessing the quality of a diagnostic test evaluation. J Gen Int Med 1989; 4:288-295.

47. Gunnarsson LG, Amilon A, Hellstrand P, Leissner P, Philipson L. The diagnosis of carpal tunnel syndrome: sensitivity and specificity of some clinical and electrophysiological tests. J Hand Surg [Br] 1997; 22B:34-37.

48. Thornbury JR, Fryback DG, Turski PA. Disk-caused root compression in patients with MR, CT myelography, and plain CT. Spine 1993; 186:731-738.

49. Schulzer M. Diagnostic tests: a statistical review. Muscle & Nerve 1994; 17:815-819.

50. Lachs MS, Nachamkin I, Edelstein PH, Goldman J, Feinstein AR, Schwartz JS. Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. Ann Int Med 1992; 117:135-140.

51. Sheps SB, Schechter MT. The assessment of Diagnostic tests. JAMA 1984; 252:2418-2422.

52. van den Hoogen HMM, Koes BW, van Eijk JTM, Bouter LM. On the accuracy of history, physical examination, and erythrocyte sedimentation rate in diagnosing low back pain in general practice. Spine 1995; 20:318-327.

53. Begg C, Cho M, Eastwood S. Improving the quality of reporting randomized controlled trials: the CONSORT statement. JAMA 1996; 276:637-639.

54. Dujardin B, Van den Ende J, Gompel AV, Unger JP, Van der Stuyft. Likelihood ratios: a real improvement for clinical decision making? Eur J Epidemiol 1994; 10:29-36.

55. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 1991; 44:763-770.

56. Simel DL, Feussner JR, Delong ER, Matchar DB. Intermediate, indeterminate, and uninterpretable diagnostic test results. Med Decis Making 1987; 7:107-114.

57. Fritz JM, Erhard RE, Delitto A, Welch WC, Nowakowski PE. Preliminary results of the use of a two-stage treadmill test as a clinical diagnostic tool in the differential diagnosis of lumbar spinal stenosis. J Spinal Disord 1997; 10:410-415.

58. Fagan TJ. Nomogram for Bayes's theorem. N Eng J Med 1975; 293:257

59. Engel GL. The need for a new medical model: a challenge for biomedicine. Science 1977; 196:129-136.

60. Taylor FK. The medical model of the disease concept. B J Psych 1976; 128:588-594.

61. Waddell G, Main CJ. Assessment of severity in low-back disorders. Spine 1984; 17:617-628.

62. Deyo RA, Loeser JD, Bigos SJ. Herniated lumbar intervertebral disk. Ann Int Med 1990; 112:598-603.

63. Ettinger B, Black DM, Nevitt MC. Contribution of vertebral deformities to chronic back pain and disability. J Bone Miner Res 1992; 7:449-455.

64. Dimitru D. Electrodiagnotic Medicine. Philadelphia, PA: Hanley & Belfus, Inc., 1995.

65. Stedman's Medical Dictionary. Ny, 1975.

66. Stevens J. AAEM minimonograph #26 The electrodiagnosis of carpal tunnel syndrome. Muscle & Nerve 1997; 20:1477-1486.

67. Jablecki CK, Andary MT, So YT, Wilkins DE, Williams FH. Literature review of the usefulness of nerve conduction studies and electromyography for the evaluation of patients with carpal tunnel syndrome. Muscle & Nerve 1993; 16:1392-1414.

68. Buchthal F, Rosenfalck A. Spontaneous electrical activity of human muscle. Electroencephalogr Clin Neurophysiol 1966; 20:321

69. Denny-Brown D, Foley JM. Fibrillations and fasiculation in voluntary muscle. Brain 1938; 61:311

70. van den Bent MJ, Oosting J, Laman DML, van Duijn H. EMG before and after cervical anterior discectomy. Acta Neurol Scand 1995; 92:332-336.

71. Negrin P, Lelli S, Fardin P. Contribution of electroymyography to the diagnosis, treatment, and prognosis of cervical disc disease: a study of 114 patients. Electromyogr Clin Neurophysiol 1991; 31:173-179.

72. Schott K, Koenig E. T-wave response in cervical root lesions. Acta Neurol Scand 1991; 84:273-276.

73. Shea P.A., Woods W.W., Werden D.H. Electromyography in diagnosis of nerve root compression syndrome. Arch Neurol Psychiat 1950; 64:93-104.

74. Tackmann W, Radu EW. Observations on the application of electrophysiological methods in the diagnosis of cervical root compressions. Eur J Neurol 1983; 22:397-404.

75. Partanen J, Partanen K, Oikarinen H, Niemitukia L, Hernesniemi J. Preoperative electroneuromyography and myelography in cervical root compression. Electromyogr Clin Neurophysiol 1991; 31:21-26.

76. Haldeman S. The electrodiagnostic evaluation of nerve root function. Spine 1984; 9:42-48.

77. Young A., Getty J., Jackson A., Kirwan E., Sullivan M., Christopher W.P. Variations in the pattern of muscle innervation by the L5 and S1 nerve roots. Spine 1983; 8:616-624.

78. Knuttson B. Comparative value of electromyographic, myelographic, and clinical-neurologic examinations in diagnosis of lumbar root compression syndromes. Acta Orthop Scand (Suppl) 1961; 49:1-135.

79. Boden SD, McCowin PR, Davis DO, Dina TS, Mark AS, Wiesel S. Abnormal magnetic-resonance scans of the cervical spine in asymptomatic subjects. JBJS [A] 1990; 72A:1178-1184.

80. Powell MC, Wilson M, Szypryt P. Prevalence of lumbar disc degeneration observed by magnetic resonance in symptomless women. Lancet 1986; December:1366-1367.

81. Weinreb JC, Wolbarsht LB, Cohen J. Prevalence of lumbosacral intervertebral disk abnormalities on MR images in pregnant and asymptomatic women. Radiology 1989; 170:125-128.

82. Jensen M, Bran-Zawadzki M, Obuchowski N. Magnetic resonance imaging of the lumbar spine in people without back pain. N Eng J Med 1994; 331:69-73.

83. Boden SD. Davis DO, Dina TS, Mark AS, Wiesel S. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. JBJS [A] 1990; 331:403-408.

84. Wilbourn A, Aminoff M.J. AAEE mini monograph #32: The electrophysiological examination in patients with radiculopathies. Muscle & Nerve 1988; 11:1011-1014.

85. Daniels DL, Grogan JP, Johansen JG, Meyer GA, Williams AL, Haughton VM. Cervical radiculopathy: computed tomography and myelography compared. Radiology 1981; 151:109-113.

86. Flax H.J., Berrios R., Rivera R. Electromyography in the diagnosis of herniated lumbar disc. Arch Phys Med Rehabil 1964; 520-524.

87. Tonzola R, Ackil AA, Shahani BT, Young RR. Usefulness of electrophysiological studies in the diagnosis of lumbsacral root disease. Ann Neurol 1981; 9:305-308.

88. Aminoff M.J., Goodin D.S., Parry G.J., Barbaro N.M., Weinstein P.R., Rosenblum M.L. Electrophysiologic evaluation of lumbosacral radiculopathies: Electromyography, late responses, and somatosensory evoked potentials. Neurology 1985; 35:1514-1518.

89. JaJoie WJ. Nerve root compression: correlation of electromyographic, myelographic, and surgical findings. Arch Phys Med Rehabil 1972; 53:390-392.

90. Wise CS, Ardizzone J. Electromyography in intervertebral disc protrusions. Arch Phys Med Rehabil 1954; 35:442-446.

91. Serra AI, Tugnoli V, Cristofori MC, Migliore A, Roccella P, Rosati G. Electrophysiological findings in patients with lumbar disc prolapse. Electromyogr Clin Neurophysiol 1984; 24:313-320.

92. Stratford P, Binkley JM, Solomon P. Assessing change over time in patients with low back pain. Phys Ther 1994; 74:528-533.

93. Leblhuber F, Reisecker F, Boehm-Jurkovic H, itzmann A, Deisenhammer E. Diagnostic value of different electrophysiologic tests in cervical disk prolapse. Neurology 1988; 38:1879-1882.

94. Parkkola R, Rytokoski U, Kormano M. Magnetic resonance imaging of the discs and trunk muscles in patients with chronic low back pain and healthy control subjects. Spine 1993; 18:830-835.

95. Wiesel SW, Tsourmas N, et.al. A study of computer assisted tomography. I. The incidence of positive CAT scans in an asymptomatic group of patients. Spine 1984; 9:549

96. British Association of Physical Medicine. Pain in the neck and arm. A multicentre trial of the effects of physiotherapy. BMJ 1966; i:253-258.

97. Rothman RH, Marvel JP. The acute cervical disk. Clin Orthopedics 1975; 109:59-68.

98. Saal JS, Saal JA, Yurth EF. Nonoperative management of herniated cervical intervertebral disc with radiculopathy. Spine 1996; 21:1877-1883.

99. Persson LCG, Carlsson CA, Carlsson JY. Long-lasting cervical radicular pain managed with surgery, physiotherapy, or a cervical collar: a prospective, randomized study. Spine 1997; 22:751-758.

100. Ellengerg M, Reina N, Ross M, Chodoroff G, Honet JC, Schwartz M, et al. Prospective evaluation of the course of disc herniations in patients with proven radiculopathy. Arch Phys Med Rehabil 1993; 74:3-8.

101. Brenner H, Savitz D. The effects of sensitivity and specificity of case selection on validity, sample size, precision, and power in hospital-based case-control series. Am J Epidemiol 1990; 132:181-192.

102. Shahani BT. The utility of proximal nerve conduction in radiculopathies: the pros. Electroencephalogr Clin Neurophysiol 1991; 78:168-170.

103. Eisen A. The utility of proximal nerve conduction in radiculopathies: the cons. Electroencephalogr Clin Neurophysiol 1991; 78:171-172.

104. Tullberg T, Svanborg E, Isacsson J, Grane P. A preoperative and postoperative study of the acuracy and value of electrodiagnosis in patients with lumbosacral disc herniation. Spine 1993; 18:837-842.

105. Pease WS, Cannell CD, Johnson EW. Median to radial latency difference test in mild carpal carpal tunnel syndrome. Muscle & Nerve 1989; 12:905-909.

106. Simpson JA. Electrical signs in the diagnosis of carpal tunnel and related syndromes. J Neurol Neurosurg Psychiatry 1956; 19:275-280.

107. Thomas JE, Lambert EH, Cseuz KA. Electrodiagnostic aspects of the carpal tunnel syndrome. Arch Neurol 1967; 16:635-641.

108. Thomas PK. Motor nerve conduction in the carpal tunnel syndrome. Neurology 1960; 10:1045-1050.

109. Gilliatt RW, Sears TA. Sensory nerve action potentials in patients with peipheral nerve lesions. J Neurol Neurosurg Psychiatry 1958; 21:109-118.

110. Brown WF, Ferguson GG, Jones MW, Yates SK. The location of conduction abnormalities in human entrapment neuropathies. Can J Neurol Science 1976; 3:111-122.

111. Fullerton PM. The effect of ischaemia on nerve conduction in the carpal tunnel syndrome. J Neurol Neurosurg Psychiatry 1963; 385-397.

112. Jablecki CK. Practice parameter for electrodiagnostic studies in carpal tunnel syndrome. Neurology 1993; 43:2404-2405.

113. Padua L, Monaco ML, Valente EM, Tonali PA. A useful electrophysiologic parameter for diagnosis of carpal tunnel syndrome. Muscle & Nerve 1996; 19:48-53.

114. Kimura J, Ayyar DR. The carpal tunnel syndrome: electrophysiological aspects of 639 symptomatic extremities. Electroencephalography and Clinical Neurophysiology 1985; 25:151-164.

115. Jackson D, Clifford JC. Electrodiagnosis of mild carpal tunnel syndrome. Arch Phys Med Rehabil 1989; 70:199-204.

116. Carroll G. Comparison of median and radial nerve sensory latencies in the electrophysiological diagnosis of carpal tunnel syndrome. Electroencephalogr Clin Neurophysiol 1987; 68:101-106.

117. Kimura J. The carpal tunnel syndrome: localization of conduction abnormalities within the distal segment of the median nerve. Brain 1979; 102:635

118. Grundberg AB. Carpal tunnel decompression in spite of normal electromyography. J Hand Surg [A] 1983; 8:348-349.

119. Redmond MD, Rivner MH. False positive electrodiagnostic tests in carpal tunnel syndrome. Muscle & Nerve 1988; 11:511-517.

120. Richman JA, Gelberman RH, Rydevik BL, et.al. Carpal tunnel volume determination by magnetic resonance imaging three-dimensional reconstruction. J Hand Surg [A] 1987; 12A:712-717.

121. Kulick MI, Ordillo G, Javidi T, Kilgor ES, Newmeyer WL. Long-term analysis of patients having surgical treatment for carpal tunnel syndrome. J Hand Surg [A] 1986; 11A:59-66.

122. Nolan WB, Alkaitis D, Glickel SZ, Snow S. Results of treatment of severe carpal tunnel syndrome. J Hand Surg [A] 1992; 17A:1020-1023.

123. Finestone HM, Woodbury GM, Collavini T, Marchuk Y, Maryniak O. Severe carpal tunnel syndrome:clinical and electrodiagnostic outcome of surgical and conservative treatment. Muscle & Nerve 1996; 19:237-239.

124. Glowacki KA, Breen CJ, Sachar K, Weiss APC. Electrodiagnostic testing and carpal tunnel release outcome. J Hand Surg [A] 1996; 21A:117-122.

125. Concannon MJ, Gainor B, Petroski GF, Puckett CL. The predictive value of electrodiagnostic studies in carpal tunnel syndrome. Plast Reconstr Surg 1997; 100:1452-1458.

126. Braun RM, Jackson WJ. Electrical studies as a prognostic factor in the surgical treatment of carpal tunnel syndrome. J Hand Surg [A] 1994; 19A:893-900.

127. Padua L, LoMonaco M, Aulisa L, Tamburrelli F, Valente EM, Padua R, et al. Surgical prognosis in carpal tunnel syndrome: usefulness of a preoperative neurophysiological assessment. Acta Neurol Scand 1996; 94:343-346.

128. Cobb TK, Amadio PC, Leatherwood DF, Schleck CD, Ilstrup DM. Outcome of reoperation for carpal tunnel syndrome. J Hand Surg [A] 1996; 21A:347-356.

129. Chang B, Dellon AL. Surgical management of recurrent carpal tunnel syndrome. J Hand Surg [Br] 1993; 18B:467-470.

130. Jablecki CK. Practice parameter for carpal tunnel syndrome. Neurology 1993; 43:2406-2409.

131. Nagi S. Disability concepts revisited: implications for prevention. In: Pope AM, Tarlov AR, editors. Disability in America: Toward a National Agenda for Prevention. Washington DC: National Academy Press, 1991:

132. Kaplan RM, Atkins CJ, Timms R. Validity of a quality of well-being scale as an outcome measure in chronic obstructive pulmonary disease. J Chron Dis 1984; 37:85

133. Croog SH, Levine S, Testa MA. The effects of antihypertensive therapy on the quality of life. N Eng J Med 1986; 314:1657

134. Million R, Hall W, Nilson KH. Assessment of the progress of the back-pain patient. Spine 1982; 7:204

135. Dwosh IL, Giles AR, Ford PM. Plasmapheresis therapy in rheumatoid arthritis: a controlled, double-blind, cross-over trial. N Eng J Med 1983; 308:1124

136. Franciosa JA, Cohn JN. Effect of isosorbide dinitrate on response to submaximal and maximal exercise in patients with congestive heart failure. Am Cardiol 1979; 43:

137. Goodwill CJ. The carpal tunnel syndrom: long term follow-up showing relation of latency measurements to response to treatment. Ann Phys Med 1965; 8:12-21.

138. Melvin JL, John, Duran R. Electrodiagnosis after surgery for the carpal tunnel syndrome. Arch Phys Med Rehabil 1968; 49:502-507.

139. Whyte J. Toward a methodology for rehabilitation research. Rehabil Res 1994; 73:428-435.

140. Deyo RA, Patrick DL. Barriers to the use of health status measures in clinical investigation, patient care, and policy research. Med Care 1989; 27:S254-S268

141. Wilson RW, Gieck JH, Gansneder BM, Perrin DH, Saliba EN, McCue FC. Reliability and responsiveness of disablement measures following acute ankle sprains among athletes. J Orthop Sports Phys Ther 1998; 27:348-355.

142. Kirshner B. Guyatt G. A methodological framework for assessing health indices. J Chron Dis 1985; 38:27-36.

143. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimally clinically important difference. Control Clin Trials1989; 10:407-415.

144. Stratford PW, Binkley JM. Applying the results of self-report measures to individual patients: an example using the Roland-Morris Questionnaire. J Orthop Sports Phys Ther 1999; 29:232-239.

145. Stratford P, Binkley JM, Riddle DL. Health status measures: Strategies and analytic methods for assessing change scores. Phys Ther 1996; 76:1109-1123.

146. Beurskens AJHM, de Vet HCW, Koke AJA. Resonsiveness of functional status in low back pain: a comparison of different instruments. Pain 1996; 65:71-76.

147. Norman GR, Stratford P, Regehr G. Methodological problems in hte retrospective computation of responsiveness to change: the lesson of Chronbach. J Clin Epidemiol 1997; 50:869-879.

148. Kaplan SJ, Glickel SZ, Eaton RG. Predictive factors in the non-surgical treatment of carpal tunnel syndrome. J Hand Surg [Br] 1990; 15B:106-108.

149. Wintman BI, Winters SC, Gelberman RH, Katz JN. Carpal tunnel release: correlations with preoperative symptomatology. Clin Orthop 1996; 326:135-145.

150. Rothstein JM, Roy SH, Wolf SL. The Rehabilitation Specialist's Handbook. Philadelphia: F.A. Davis, 1991.

151. Porterfield JA, DeRosa C. Principles of Mechanical Neck Disorders. Philadelphia: W. B. Saunders Company, 1995.

152. Michelsen JJ, Mixter WJ. Pain and disability of shoulder and arm due to herniation of the enucleus pulposus of cervical intervetebral disks. N Eng J Med 1944; 231:279-287.

153. Fager CA. Identification and management of radiculopathy. Neurosurg Clin 1993; 4:1-12.

154. Viikari-Juntura E. Interexaminer reliability of observations in physical examinations of the neck. Phys Ther 1987; 67:1526-1532.

155. Silver M, McElroy A, Morrow L, Heafner BK. Further standardization of manual muscle test for clinical study: applied in chronic renal disease. Phys Ther 1970; 50:1456-1465.

156. Wadsworth CT, Krishnan R, Sear M, Harrold J, Nielsen DH. Intrarater reliability of manual muscle testing and hand-held dynametric testing. Phys Ther 1987; 67:1342-1347.

157. Phalen GS. The carpal tunnel syndrome: seventeen years experience in diagnosis and treatment of six hundred and fifty four hands. JBJS [A] 1966; 48A:211-228.

158. Jackson DA, Clifford JC. Electrodiagnosis of mild carpal tunnel syndrome. Arch Phys Med Rehabil 1989; 70:199-204.

159. Loong SC. The carpal tunnel syndrome: a clinical electrophysiologic study in 250 patients. Clin Exper Neurol 1977; 14:51-65.

160. De Krom MC, Knipschild PG, Kester AD. Efficacy of provocative tests for diagnosis of carpal tunnel syndrome. Lancet 1990; 335:393-395.

161. Shivde AJ, Fisher MA. The carpal tunnel syndrome: a clinical electrodiagnostic analysis. Electromyogr Clin Neurophysiol 1981; 21:143-153.

162. Stevens JC, Son S, Beard CM, O'Fallon WM, Kulland KT. Carpal tunnel syndrome in Rochester. Minnesota 1961 to 1980. Neurology 1988; 38:134-138.

163. Borg K, Lindblom U. Increase of vibration threshold during wrist flexion in patients with carpal tunnel syndrome. Pain 1986; 26:211-219.

164. Golding DN, Rose DM, Selvarajah K. Clinical tests for carpal tunnel syndrome. B J Rheumatol 1986; 25:388-390.

165. Youdas JW, Carey JR, Garrett TR. Reliability of measurements of cervical spine range of motion---comparison of three methods. Phys Ther 1991; 71:98-106.

166. Lowery WD, Horn TJ, Boden SD, Wiesel SW. Impairment evaluation based on spinal range-of-motion in normal subjects. J Spinal Disord 1992; 5:398-402.

167. Erhard, R. Manual therapy in the cervical spine. Continuing Education Series. American Physical Therapy Association, 1996; 3-3.

168. Cyriax J. Orthopaedic Medicine. 8 ed. London: Bailliere Tindall, 1999.

169. Maitland GD. Cervical Spine. In: Maitland GD, editor. Vertebral manipulation. 5th ed. London: Butterworth & Co., 1986:

170. Hole DE, Cook JM, Bolton JE. Reliability and concurrent validity of two instruments for measuring cervical range-of-motion: effects of age and gender. Manual Therapy 1995; 1:36-42.

171. Ordway NR, Seymour R, Donelson RG, Hojnowski L, Lee E, Edwards WT. Cervical sagittal range-of-motion analysis using three methods: cervical range-of-motion device, 3 space, and radiography. Spine 1997; 22:501-508.

172. Kuhlman KA, Hennessey WJ. Sensitivity and specificity of carpal tunnel syndrome signs. Am J Phys Med Rehabil 1997; 76:451-457.

173. Radecki P. A gender specific wrist ratio and the likelihood of a median nerve abnormality at the carpal tunnel. Am J Phys Med Rehabil 1994; 73:157-162.

174. Johnson EW, Gatens T, Poindexter D, Bowers D. Wrist dimensions: correlation with median sensory latencies. Arch Phys Med Rehabil 1983; 64:556-557.

175. Gordon C, Johnson EW, Gatens PF, Ashton JJ. Wrist ratio correlation with carpal tunnel syndrome in industry. Am J Phys Med Rehabil 1988; 67:270-272.

176. Sposato RC, Riley MW, Ballard JL, Stentz TL, Glismann CL. Wrist squareness and median nerve impairment. J Occup Environ Med 1995; 37:1122-1126.

177. Butler DS. Mobilization of the Nervous System. New York: Churchhill Livingstone, 1991.

178. Gelberman RH, Hargens AR, Lundborg GN, Akeson WH. The carpal tunnel syndrome: a study of carpal canal pressure. JBJS [A] 1981; 63A:380-383.

179. Lundborg GN, Gelberman RH, Minteer-Convery M, Lee YF, Hargens AR. Median nerve compression in the carpal tunnel: functional response to experimentally induced controlled pressure. J Hand Surg [A] 1982; 7:252-259.

180. Muhle C, Bischoff L, Weinert D. Exacerbated pain in cervical radiculopathy at axial rotation, flexion, extension, and coupled motions of the cervical spine: evaluation by kinematic magnetic resonance imaging. Invest Radiol 1998; 33:279-288.

181. Howe JF, Loeser JD, Calvin WH. Mechanosensitivity of dorsal root ganglia and chronically injured axons: a physiological basis for the radicular pain of nerve root compression. Pain 1977; 3:25-41.

182. Rappaport ZH, Dever M. Experimental pathophysiological correlates of clinical symptomatology in peripheral neuropathic pain syndromes. Stereotact Funct Neurosurg 1990; 54 & 55:90-95.

183. Spurling RG, Scoville WB. Lateral rupture of the cervical intervrtebral discs: a common cause of shoulder and arm pain. Surg Gynecol Obstet 1944; 78:350-358.

184. Davidson RI, Dunn EJ, Metzmaker JN. The shoulder abduction relief test in the diagnosis of radicular pain in cervical extradural compressive monoradiculopathies. Spine 1981; 6:441-446.

185. Elvey RL. The Investigation of arm pain: signs of adverse responses to the physical examination of the brachial plexus and related tissues. In: Boyling JD, Palastanga N, editors. Grieve's Modern Manual Therapy. 2nd Ed. ed. New York: Churchill Livingstone, 1994:577-585.

186. Durkan JA. A new diagnostic test for carpal tunnel syndrome. JBJS [A] 1991; 13A:535-538.

187. Paley D, McMurtry RY. Median nerve compression test in carpal tunnel syndrome diagnosis reproduces signs and symptoms in affected wrists. Orthop Rev 1985; 15:41-45.

188. Tinel J. Le signe du fourmillement dans les lesions des nerfs peripheriques. La Presse Medicale 1915; 47:388-399.

189. Monsivais JJ, Sun Y. Tinel's sign or percussion test? Developing a better method of evoking a Tinel's sign. J Southern Orthop Assoc 1997; 6:186-189.

190. Sandmark H, Nisell R. Validity of five common manual neck pain provoking tests. Scand Rehabil Med 1995; 27:131-136.

191. Beatty RM, Fowler FD, Hanson EJ. The abducted arm as a sign of ruptured cervical disc. Neurosurgery 1987; 21:731-732.

192. Roberts DK, MacKay KA. Microhemorrahagic maculopathy associated with aerobic exercise. J M Optom Assoc 1987; 58:415-418.

193 Martini B, Buja GF, Canciani B, Maddalena F, Nava A. Atrial desynchronization induced by Valsalva's maneuver in a patient with reciprocating supraventricular tachycardia and Wolff-Parkinson-White syndrome. Giornale Italiano di Cardiologia 1987; 17:830-833.

194. Schartum S. Ventricular arrest caused by the Valsalva maneuver in a patient with Adams-Stokes attacks accompanying defecation. Acta Medica Scandinavica 1968; 184:65-68.

195. Elvey RL. editor. Painful restriction of shoulder movement: a clinical observation study. Western Australia Institute of Technology, Perth. Australia 1979.

196. Dellon AL. Tinel or not to Tinel. J Hand Surg [Br] 1988; 12B:216-216.

197. Skie M, Nabil AE, Jackson WT. Carpal tunnel changes during wrist flexion and extension seen by magnetic resonance imaging. J Hand Surg [A] 1990; 15A:934-939.

198. Kaplan SJ, Glickel S, Eaton RG. Predicitive factors in the non-surgical treatment of carpal tunnel syndrome. J Hand Surg [Br] 1990; 15B:106-108.

199. De Smet L, Steenwerckx A, Van Den Bogaert G, Cnudde P, Fabry G. Value of clinical provocative tests in carpal tunnel syndrome. Acta Orthopaedica Beligica 1995; 61:177-182.

200. Gonzalez Del Pino J, Delgado-MartinezAD, Gonzalez-Gonzalez I, Lovic A. Value of the carpal compression test in the diagnosis of carpal tunnel syndrome. J Hand Surg [Br] 1997; 22 B:38-41.

201. Mossman SS. Blau JN. Tinel's sign and the carpal tunnel syndrome. BMJ 1987; 294:680

202. Seror P. Phalen's test in the diagnosis of carpal tunnel syndrome. J Hand Surg 1988; 13B:383-385.

203. LaBan MM, Friedman NA, Zemenick GA. "Tethered" median nerve stress test in chronic carpal tunnel syndrome. Arch Phys Med Rehabil 1988; 67:803-804.

204. Gellman H, Gelberman RH, Tan AM, Botte MJ. Carpal tunnel syndrome: an evaluation of the provocative diagnostic tests. JBJS [A] 1986; 68A:735-737.

205. Heller L, Ring H, Costeff H, Solzi P. Evaluation of Tinel's and Phalen's signs in diagnosis of the carpal tunnel syndrome. Eur Neurol 1986; 25:40-42.

206. Kuschner SH. Ebramzadeh E, Johnson D, Brian WW. Tinel's sign and Phalen's test in carpal tunnel syndrome. Orthopaedics 1992; 15:1297-1302.

207. Stewart JD. Eisen A. Tinel's sign and the carpal tunnel syndrome. BMJ 1978; 2:(1125)1126

208. Rietz KA. Onne L. Analysis of sixty-five operated cases of carpal tunnel syndrome. Acta Chir Scand 1967; 133:443-447.

209. Tetro MA, Evanoff BA, Hollstien SB, Gelberman RH. A new provocative test for carpal tunnel syndrome: assessment of wrist flexion and nerve compression. JBJS [Br] 1998; 80B:(3)493-498.

210. Williams TM. Mackinnon SE, Novak CB. McCabe S, Kelly L. Verification of the pressure provocative test in carpal tunnel syndrome. Ann Plast Surg 1992; 29:(1)8-11.

211. Novak CB. Mackinnon SE, Brownlee R, Kelly L. Provocative sensory testing in carpal tunnel syndrome. JBJS [Br] 1992; 17B:(2)204-208.

212. Marx RG. Hudak PL, Bombardier C, Graham B, Goldsmith C, Wright JG. The reliability of physical examination for carpal tunnel syndrome. J Hand Surg [Br] 1998; 23B:499-502.

213. Szabo RM. Slater RR, Farver TB. Stanton DB, Sharman WK. The value of diagnostic testing in carpal tunnel syndrome. J Hand Surg [A] 1999; 24A:704-714.

214. Gelmers HG. The significance of Tinel's sign in the diagnosis of carpal tunnel syndrome. Acta Neurochirugica 1979; 49:255-258.

215. Durkan JA. The carpal-compression test: An instrumented device for diagnosing carpal tunnel syndrome. Orthop Rev 1994; 522-525.

216. Wainner RS, Boninger ML, Balu G, Burdett R, Helkowski W. The Durkan gauge and Carpal Compression Test: Instrument Validity and Diagnostic Test Properties in a Case Series of 57 Consecutively Tested Hands. J Orthop Sports Phys Ther (accepted for publication, 2000).

217. Yii N, Elliot D. A study of the dynamic relationship of the lumbrical muscles and the carpal tunnel. J Hand Surg [Br] 1994; 19B:439-443.

218. Seror P. Tinel's sign in the diagnosis of carpal tunnel syndrome. J Hand Surg [Br] 1987; 12B:364-365.

219. Steiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford: Oxford University Press, 1995.

220. Vernon H, Mior S. The neck disability index: a study of reliability and validity. J Manip Physiol Ther 1991; 14:407-415.

221. Levine DW, Simmons BP, Koris MJ, Daltroy LH, Hohl GG, Fossel AH, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. JBJS [A] 1993; 75-A:1585-1592.

222. Waddell G, Newton M, Henderson I, Somerville D, Main C. A Fear-avoidance beliefs questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993; 52:157-168.

223. Katz JN, Stirrat CR. A self-administered hand diagram for the diagnosis of carpal tunnel syndrome. J Hand Surg [A] 1990; 15A:360-363.

224. Delitto A. Are measures of function and disability important in low back care? Phys Ther 1994; 74:452-462.

225. Riddle DL, Stratford PW. Use of Generic versus region specific functional status measures on patients with cervical spine disorders. Phys Ther 1998; 78:951-963.

226. Westaway MD, Stratford PW, Binkley JM. The patient-specific functional scale: validation of its use in persons with neck dysfunction. J Orthop Sports Phys Ther 1998; 27:-331

227. Katz JN, Gelberman RH, Wright EA, Lew RA, Liang MH. Responsiveness of self-reported and objective measures of disease severity in carpal tunnel syndrome. Med Care 1994; 32:1127-1133.

228. Amadio PC, Silverstein MD, Ilstrup DM, Schleck CD, Jensen LM. Outcome assessment for carpal tunnel surgery: the relative responsiveness of generic, arthritis-specific, disease specific, and physical examiniation measures. JBJS [A] 1996; 21A:338-346.

229. Katz JN, Punnett L, Simmons BP, Fossel AH, Mooney N, Keller RB. Workers' compensation recipients with carpal tunnel syndrome: the validity of self-reported health measures. Am J Public Health 1996; 86:52-56.

230. Katz JN, Stirrat CR, Larson MG, Fossel AH, Eaton HM, Liang MH. A self-administered hand symptom diagram for the diagnosis and epidemiologic study of carpal tunnel syndrome. J Rheumatol 1990; 17:1494-1498.

231. Bessette L, Keller RB, Lew RA, Simmons BP, Fossel AH, Mooney N, et al. Prognostic value of a hand symptom diagram in surgery for carpal tunnel syndrome. J Rheumatol 1997; 24:726-734.

232. Ransford AO, Douglas C, Mooney V. The pain drawing as an aid to the psychologic evaluation of patients with low-back pain. Spine 1976; 1:127-134.

233. Chan CW, Goldman S, Ilstrup DM, Kunselman AR, O'Neill PI. The pain drawing and Waddell's nonorganic physical signs in chronic low-back pain. Spine 1993; 18:1717-1722.

234. Downie WW, Leatham PA, Rhind VM. Studies with pain rating scales. Ann Rheum Dis 1978; 37:378-381.

235. Wilkie D, Lovejoy N, Dodd M, Tesler M. Cancer pain intensity measurement; concurrent validity of three tools- finger dynameter, pain intensity number scale, visual analogue scale. Hospice Journal 1990; 6:1-13.

236. Langley GB, Sheps SB. The visual analogue scale: Its use in pain measurement. Rheumatol Int 1985; 5:145-148.

237. Scott J, Huskisson EC. Vertical or horizontal visual analogue scales. Ann Rheum Dis 1979; 38:560

238. Letham J, Slade PD, Troup JD, Bentley G. Outline of a Fear-Avoidance Model of exaggerated pain perception--I. Behav Res Ther 1983; 21:401-408.

239. Slade PD, Troup JD, Letham J, Bentley G. The Fear-Avoidance Model of exaggerated pain perception--II. Behav Res Ther 1983; 21:409-416.

240. Klenerman L, Slade PD, Stanley IM. The prediction of chronicity in patients with an acute attack of low back pain in a general practice setting. Spine 1995; 20:478-484.

241. Fritz JM. The efficacy of a homogenous treatment approach versus a classification approach to the treatment of work-related low back pain. A randomized trial. Doctoral Dissertation, 1998.

242. Fleiss JL. The Measurement of Interrater Agreement. In: Anonymous Statistical Methods for Rates and Proportions. 2nd Ed ed. New York: John Wiley & Sons, 1981:213-235.

243. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. Norwalk, CT: Appleton & Lange, 1993.

244. McQuillen MP GF. Serial ulnar nerve conduction velocity measurements in normal subjects. J Neurol Neurosurg Psych 1969; 32:144-148.

245. Chaudry V, Cornblath DR, Avila O, Freimer ML, Glass JD, Reim J, et al. Inter and intra-examiner reliability of nerve conduction measurements in normal subjects. Ann Neurol 1991; 30:841-843.

246. McQuillen MP, Gorin FJ. Serial ulnar nerve conduction velocity measurements in normal subjects. J Neurol Neurosurg Psych 1969; 32:144-148.

247. Mitchell SK. Interobserver agreement, reliability, and generalizability of data collected in observational studies. Psycological Bulletin 1979; 86:376-390.

248. Moore J. Reliability of median nerve conduction measurements. 1997. (GENERIC) Ref Type: Unpublished Work

249. Khatri B., Baruah J., McQuillen M.P. Correlation of Electromyography with computed tomography in evaluation of lower back pain. Arch Neurol 1984; 41:594-597.

250. Jabre J. Surface recording of the H-reflex of the flexor carpi radialis. Muscle & Nerve 1981; Sept/Oct:435-438.

251. Kendall FP, McCreary EK. Muscles, Testing and Function. 3rd Ed. ed. Baltimore: 1983.

252. Rothstein JM. Measurement and Clinical Practice: Theory and Application. In: Rothstein JM, editor. Measurement in Physical Therapy. New York: Churchill Livingstone, 1985:1-46.

253. McHorney CA, Ware JE, Raczek AE. The MOS 36-item short form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993; 31:247-263.

254. Strube MJ, Delitto A. Reliability and Measurement Theory. In: Craik RL, Oatis CA, editors. Gait Analysis. Theory and Application. St. Louis, MO: Mosby-Year Book, Inc., 1995:88-111.

255. Fleiss JL. Reliability of Measurement. In: Anonymous The Design and Analysis of Clinical Experiments. NY, New York: Wiley & Sons, 1986:1-32.

256. Strube MJ. Interpreting the NIOSH Analyses. 1989. Govt. Publication

257. Bartko JJ, Carpenter WT. On the methods and theory of reliability. J Nerv Men Dis 1976; 163:307-317.

258. Haley SM, Osberg JS. Kappa coefficient calculation using multiple ratings per subject: a special communications. Phys Ther 1989; 69:970-974.

259. Cohen J. A coefficient of agreement for nominal scales. Ed Psych Meas 1960; 20:37-46.

260. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159-174.

261. Denegar CR, Ball DW. Assessing reliability and precision of measurement: an introduction to intraclass correlation and standard error of measurement. J Sports Rehabil 1993; 2:35-42.

262. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psych Bull 1979; 86:420-428.

263. Morgan G, Wilbourn AJ. Cervical radiculopathy and coexisting distal entrapment neuropathies: Double crush syndromes? Neurology 1989; 50:78-83.

264. Wilbourn A. Double crush syndrom: a critical analysis. Neurology 1997; 49:21-29.

265. Murphy JM, Berwick DM, Weinstein MC, Borus JF, Budman SH, Klerman GL. Performance of screening and diagnostic tests: application of recever operating characteristic analysis. Arch Gen Psych 1987; 44:550-555.

266. Hanley JA. McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143:29-36.

267. Galen RS. Gambino SR. Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses. New York: John Wiley and Sons, 1975.

268. Morise AP. Duval RD. Comparison of three Bayesian methods to estimate postest probability in patients undergoing exercise stress testing. Am Cardiol 1989; 64:1117-1122.

269. Kraemer C TS. Statistical Power Analysis in Research. Beverly Hills, CA. Sage, 1987. Beverly Hills. CA: Sage. 1987.

270. Shrout PE. Spitzer RL, Fleiss JL. Quantification of agreement in psychiatric diagnosis revisited. Arch Gen Psych 1987; 44:172-177.

271. Simel DL. Likelihood ratio calculations. Personal Communication: 2000.

272. Hosmer DW. Lemeshow S. Applied Logistic Regression. New York, NY. John Wiley & Sons. 1989:142-144.

273. Freedman DA. A note on screening regression equations. The American Statistician 1983; 37:152-155.

274. Radhakrishnan K, Litchy WJ. O'Fallon M, Kurlan LT. Epidemiology of cervical radiculopathy. A population-based study from Rochester, Minnesota, 1976 through 1990. Brain 1994; 117:325-335.

275. Atroshi I. Gummersson C, Johnsson R, Ornstein E, Ranstam J, Ingmar R. Prevalence of carpal tunnel syndrome in a general population. JAMA 1999; 282:153-158.

276. Dillin W. Booth R, Cuckler J, Balderston R, Simeone F, Rothman R. Cervical radiculopathy: A review. Spine 1986; 11:(10)988-991.

277. Sackett DL. Haynes RB. Guyett GH. Clinical Epidemiology: A Basic Science for Clinical Medicine. 2nd Ed. ed. Boston, MA: Little, Brown, and Company, Inc., 1991.

278. Fritz JM. Delitto A, Vignovic M, Busse RG. Interrater reliability of judgements of the centralization phenomenon and status change during movement testing in patients with low back pain. Arch Phys Med Rehabil 1999; 80:1-5.

279. Sampath P. Bendebba M, Davis JD, Ducker T. Outcome in patients with cervical radiculopathy. Prospective, multicenter study with independent clinical review. Spine 1999; 24:(6)591-597.

280. Cloward RB. Cervical diskography. A contribution to the etiology and mechanism of neck, shoulder, and arm pain. Ann Surg 1959; 150:1052-1064.

281. Feurerstein M. Miller VL, Burrell LM, Berger R. Occupational upper extremity disorders in the federal work force: prevalence, health care expenditures, and patterns of work disability. J Occup Environ Med 1998; 40:546-555.

282. Duncan KH. Lewis RC. Foreman KA, Nordyke MD. Treatment of carpal tunnel syndrome by members of the American Society for Surgery of the Hand: results of a questionnaire. J Hand Surg [A] 1987; 12A:384-391.

283. Pryse-Phillips W. Validation of a diagnostic sign in carpal tunnel syndrome. J Neurol 1984; 47:870-872.

284. Gerr F, Letz R. The sensitivity and specificity of tests for carpal tunnel syndrome vary with the comparison subjects. J Hand Surg [Br] 1998; 23B:151-155.

285. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chron Dis 1986; 11:897-906.

286. Burke DT, Burke MAM, Bell R, Stewart GW, Mehdi RS, Kim HJ. Subjective swelling: a new sign for carpal tunnel syndrome. Am J Phys Med Rehabil 1999; 78:504-508.

287. Bernstein J. Current concepts review: decision analysis. JBJS [A] 1997; 79A:1404-1414.

288. Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chron Dis 1987; 40:171-178.

289. Delitto A. Subjective measures and clinical decision making. Phys Ther 1989; 69:(7)585-589.

290. Feinstein AR. Clinimetrics. New Haven CT: Yale University Press, 1987.

291. Volinn E, Diehr P, Cio MA, Loeser JD. Why does geographic variation in health care practices matter: and seven questions to ask in evaluating studies on geographic variation. Spine 1994; 19:2092S-2100S.

292. Buchsbaum DG, Buchanan RG, Centor RM, Schnoll SH, Lawton MJ. Screening for alchohol abuse using CAGE scores and likelihood ratios. Ann Int Med 1991; 115:774-777.