



Diagnosis of Carpal Tunnel Syndrome
Evidence Report

Contents

I. Introduction	5
II. Methodology	5
a. Literature search	5
b. Exclusion criteria	6
c. Inclusion criteria	6
d. Rating the evidence	7
e. Data extraction	7
f. Analysis	8
g. Guideline recommendations	9
III. Results	9
IV. Limitations of the evidence base	14
References	15
Evidence Tables	
<i>Evidence Table 1.</i> Study and patient characteristics of papers comparing diagnostic tests to a reference standard of signs, symptoms of CTS and electrodiagnostic test	16
<i>Evidence Table 2.</i> Diagnostic data comparing index tests to a reference standard of signs, symptoms of CTS and electrodiagnostic test.....	18
<i>Evidence Table 3.</i> Comparing index tests to a reference standard of signs, symptoms of CTS and electrodiagnostic test.	23
<i>Evidence Table 4.</i> Study and patient characteristics of papers comparing clinical and/or electrodiagnostic tests and post-surgical outcomes	28
<i>Evidence Table 5.</i> Data characteristics of studies comparing clinical and/or electrodiagnostic tests and post-surgical outcomes	29
<i>Evidence Table 6.</i> Data comparing clinical and/or electrodiagnostic tests and post-surgical outcomes	30
<i>Evidence Table 7.</i> Data comparing clinical and/or electrodiagnostic tests and post-surgical outcomes (by severity)	33
<i>Evidence Table 8.</i> Data comparing pre- and post-surgical electrodiagnostic test results	35
<i>Evidence Table 9.</i> Study and patient characteristics of papers comparing symptoms and electrodiagnostic tests	37
<i>Evidence Table 10.</i> Data comparing symptoms and electrodiagnostic tests (1).....	41
<i>Evidence Table 11.</i> Data comparing symptoms and electrodiagnostic tests (2).....	46
<i>Evidence Table 12.</i> Study and patient characteristics of papers on splinting, steroid injection, and lifestyle change (1)	51
<i>Evidence Table 13.</i> Study and patient characteristics of papers on splinting, steroid injection, and lifestyle change (2)	52
<i>Evidence Table 14.</i> Outcomes data on steroid injections	53
<i>Evidence Table 15.</i> Outcomes data on splinting	56
<i>Evidence Table 16.</i> Outcomes data on lifestyle changes	57
<i>Evidence Table 17.</i> Outcomes data on steroid injections by electrodiagnostic tests (1)	57
<i>Evidence Table 18.</i> Outcomes data on steroid injections by electrodiagnostic tests (2)	59
<i>Evidence Table 19.</i> Outcomes data on splinting by electrodiagnostic tests	60
<i>Evidence Table 20.</i> Percent improvement due to steroid injection or splinting	61
<i>Evidence Table 21.</i> Reasons for excluding studies	61

Appendixes

Appendix A. AAOS Carpal Tunnel Syndrome Diagnosis Guideline Work Group	62
Appendix B. Levels of evidence	63
Appendix C. Data extraction forms	64
Appendix D. Bibliography of Excluded Studies	67
Appendix E. Bibliography of Included Studies	97

Figures

Figure 1. Flowchart of literature review process (Questions 1 & 2).....	105
Figure 2. Flowchart of literature review process (Question 3)	106
Figure 3. Flowchart of literature review process (Question 4)	107
Figure 4. Flowchart of literature review process (Question 5)	108
Figure 4.1 Studies on clinical tests	109
Figure 4.2 Studies on electrodiagnostic tests	109
Figure 4.3 Studies on signs/symptoms	110
Figure 4.4 Studies on median nerve electrodiagnostic tests	110
Figure 4.5 Studies on comparative electrodiagnostic tests	111
Figure 5. Level of evidence of studies addressing questions 1 & 2	112
Figure 6. Level of evidence of studies addressing question 3	112
Figure 7. Level of evidence of studies addressing question 4	113
Figure 8. Level of evidence of studies addressing question 5	113
Figure 9. Types of study design that compare CTS diagnostic tests	114
Figure 10. All tests	115
Figure 11. Case-control study design (electrodiagnostic and clinical tests)	115
Figure 12. Cohort study design (electrodiagnostic and clinical tests)	116
Figure 13. Cross sectional study design (electrodiagnostic and clinical tests)	116
Figure 14. Electrodiagnostic tests case-control study design	117
Figure 15. Clinical tests case-control study design	117
Figure 16. Clinical/Psychomotor tests	118
Figure 17. Flick, Thenar sign	118
Figure 18. Pinch strength	119
Figure 19. Provocative tests	119
Figure 20. Carpal compression tests	120
Figure 21. Phalen's test	120
Figure 22. Tinel's sign	121
Figure 23. Pressure provocative tests	121
Figure 24. Sensory tests	122
Figure 25. Touch tests	122
Figure 26. Vibrometry	123
Figure 27. Two-point discrimination, Semmes-Weinstein monofilament, ridge/gap detection.	123
Figure 28. Systematic	124
Figure 29. Median mixed nerve – ulnar mixed nerve	124
Figure 30. Ulnar comparative amplitude	125
Figure 31. Ulnar co comparative latency	125
Figure 32. Median motor – ulnar motor	126
Figure 33. Second lumbrical – Interosseous (2L-INT)	126
Figure 34. Thenar studies	127
Figure 35. Median sensory – ulnar sensory: wrist to ring finger	127
Figure 36. Sensory ulnar latency comparative	128
Figure 37. Sensory ulnar velocity comparative	128
Figure 38. Median sensory – radial sensory: wrist-to-thumb	129

Figure 39. Radial comparative velocity	129
Figure 40. Radial comparative latency	130
Figure 41. Median sensory – median mixed wrist to palm vs. forearm to digit	130
Figure 42. Segment comparison ratio	131
Figure 43. Segment comparison latency	131
Figure 44. Median motor – wrist to palm	132
Figure 45. Median motor – APB-DML	132
Figure 46. Median motor – DML amplitude, non-APB DML	133
Figure 47. Unique tests	133
Figure 48. Median sensory /median mixed – wrist to palm	134
Figure 49. Median sensory/median mixed – wrist to palm test velocity	134
Figure 50. Median sensory/median mixed – DSL	135
Figure 51. Median sensory – wrist to palm	135
Figure 52. Median sensory velocity D2 to wrist, sensory amplitude digit, sensory velocity digit.....	136
Figure 53a. Sensory latency digit 1 to wrist	136
Figure 53b. Sensory latency digit 2 to wrist	137
Figure 54. Sensory latency digit 3 to wrist	137
Figure 55. Sensory latency digit 4 to wrist	138
Figure 56. Case control studies grouped by EDT: “Comparison of Median to Ulnar Sensory nerves, wrist-to-ring finger”	139
Figure 57. Case control studies grouped by EDT: “Median motor nerve, wrist-to-palm” ...	139
Figure 58. Case control studies grouped by EDT: “Median motor nerve, wrist-to-palm measuring DML	140
Figure 59. Case control studies grouped by EDT: “Median sensory/mixed nerve, wrist-to-palm”	140
Figure 60. Case control studies grouped by EDT: “Median sensory/mixed nerve wrist-to-palm measuring DML	141
Figure 61. Case control studies grouped by EDT: “Median sensory/mixed nerve, wrist-to-palm” measuring wrist-to-palm velocity	141
Figure 62. Case control studies grouped by EDT: “Median sensory nerve, wrist-to-digit” ...	142
Figure 63. Case control studies grouped by EDT: “Comparison of median nerve, wrist-to-palm segment to other median nerve segment (i.e., elbow-to-wrist, palm-to-digit)”	142
Figure 64. Cohort studies Question 3: Relationship between symptoms and EDT	143
Figure 65. Cross sectional studies Question 3. Relationship between symptoms and EDT ...	143
Figure 66. Cross sectional studies grouped by EDT: “Combination of two or more EDT or NCS”	144
Figure 67. Cross sectional studies grouped by EDT: “Comparison of median to ulnar sensory nerves, wrist to digit”	144
Figure 68. Cross sectional studies grouped by EDT: “Median nerve inching (White et al. only).....	145
Figure 69. Cross sectional studies grouped by EDT: “Median sensory/mixed nerve, wrist-to-palm”	145
Figure 70. Cross sectional and cohort studies	146
Figure 71. Cross sectional and cohort studies grouped by EDT: “Combination of two or more EDT or NCS”	146
Figure 72. Cross sectional and cohort studies grouped by EDT: “Combination of median to ulnar sensory nerves, wrist to digit” measuring latency difference	147
Figure 73. Cross sectional and cohort studies grouped by EDT: “Median nerve inching (White et al. only).....	148

Figure 74. Cross sectional and cohort studies grouped by EDT: “Median sensory/mixed nerve, wrist-to-palm”	148
Figure 75. Cross sectional and cohort studies grouped by EDT: “Median sensory/mixed nerve, wrist-to-palm” measuring wrist-to-palm velocity	149
Figure 76. ROC space combination of two or more EDT or NCS	149
Figure 77. ROC space comparison of median to ulnar sensory nerves, wrist-to-digit	150
Figure 78. ROC space median sensory/mixed nerve, wrist-to-palm	150
Figure 79. ROC space median nerve inching	151
Figure 80a. ROC space median sensory/mixed nerve, wrist-to-palm	151
Figure 80b. ROC space median motor nerve, wrist-to-palm measuring DML	152
Figure 81. ROC space median sensory nerve, wrist-to-digit	152
Figure 82. ROC space comparison of median to radial sensory nerves, wrist-to-thumb	153
Figure 83. ROC space comparison of median nerve, wrist-to-palm segment to other median nerve segment (i.e., elbow-to-wrist, palm-to-digit)	153
Figure 84. ROC space median motor nerve TLI	154
Figure 85. Case control studies Question 3: Relationship between symptoms and EDT	154
Figure 86. Case control studies grouped by EDT: “Combination of two or more EDT/NCS ..	155
Figure 87. Case control studies grouped by EDT: “Comparison of median to ulnar mixed nerves, wrist-to-palm”	155
Figure 88. Case control studies grouped by EDT: “Comparison of median to ulnar mixed nerves, wrist-to-palm” and measuring latency difference	156
Figure 89. Case control studies grouped by EDT: “Comparison of median to ulnar motor nerves	156
Figure 90. Case control studies grouped by EDT: “Comparison of median to ulnar motor nerves” comparing 2 nd lumbrical to interosseous	157
Figure 91. Case control studies grouped by EDT: “Comparison of median to ulnar motor nerves” comparing thenar muscles	157
Figure 92. Case control studies grouped by EDT: “Comparison of median to radial sensory nerves, wrist-to-thumb”	158
Figure 93. Case control studies grouped by EDT: “Comparison of median to ulnar sensory nerves, wrist-to-digit”	158
Figure 94. Case control studies grouped by EDT: “Comparison of median to ulnar sensory nerves, wrist-to-digit” measuring latency difference	159
Figure 95. Summary of findings as reported by authors	160
Figure 96. Correlation of positive surgical outcomes with positive clinical tests	161
Figure 97. Correlation of positive surgical outcomes with positive electrodiagnostic tests	161
Figure 98. Correlation of positive surgical outcomes with positive clinical plus electrodiagnostic tests	162
Figure 99. Meta-regression of positive surgical outcomes with positive clinical plus electrodiagnostic tests	162
Figure 100. Steroid injections	163
Figure 101. Splinting data	163
Figure 102. Steroid injection DML improvement	164
Figure 103. Steroid injection and mean change in SSS scores	164
Figure 104. Magnetic resonance imaging (MRI) – Sensitivity	165
Figure 105. Magnetic resonance imaging (MRI) – Specificity	166
Figure 106. Magnetic resonance imaging (MRI) – ROC curve	167

I. Introduction

This evidence report presents the results of a systematic review of published studies on the diagnosis of carpal tunnel syndrome (CTS) in adults. It is intended to serve as a companion document to the American Academy of Orthopaedic Surgeons (AAOS) CTS Diagnosis Guideline and an information resource for decision makers and developers of practice guidelines and recommendations. It also should serve to highlight gaps in the literature and areas that require future research.

II. Methodology

The methods used to perform this systematic review were employed to minimize bias and random errors in the selection and summary of the available evidence (Mulrow et al. 1997; Cook et al. 1997; Deeks 2001; Guyatt et al. 1986). This process is vital to the development of reliable and accurate clinical recommendations for diagnosing carpal tunnel syndrome.

In order to address the most pertinent diagnostic issues, the AAOS Carpal Tunnel Syndrome Diagnosis Guideline Work Group (see [Appendix A](#)) formulated the following key questions to guide this review.

1. a. What is the diagnostic test(s) with the best performance for the diagnosis of CTS?
b. What is the diagnostic test(s) plus electrodiagnostic test(s) with the best performance for the diagnosis of CTS?
2. What is the hierarchy of electrodiagnostic tests that should be followed when confirming a CTS diagnosis?
3. What is the relationship between a patient's symptoms (including duration) and the results of electro diagnostic tests?
4. a. Is there a correlation between clinical test(s) and patient post-surgical outcomes?
b. Is there a correlation between clinical test(s) plus electrodiagnostic tests and patient post-surgical outcomes?
5. Do targeted (at the wrist) steroid injections, splinting, or activity /ergonomic changes lead to more accurate diagnosis of CTS?

A comprehensive search of the medical literature was conducted to identify the evidence available to address these five key questions. Detailed information about the studies used in the meta-analyses was abstracted, and the results are presented as evidence tables. Meta-analyses of studies of diagnostic test comparisons were conducted to provide answers to the key questions.

a. Literature search

The published literature was searched using the PubMed electronic database to identify potentially relevant studies that shed light on the questions. A manual search was performed of

the bibliographies of all publications accepted for inclusion into the evidence base. In addition, the bibliographies of recent review articles were searched for potentially relevant citations.

The PubMed search included the following search strategies, with limits of publication dates 1966 to present, English language, and humans:

Questions 1 & 2: (((*"Diagnosis"*[MeSH] OR *"diagnosis"*[Subheading] OR *diagnosis*[Text Word]) AND (*"Carpal Tunnel Syndrome"*[MeSH] OR (*"carpal tunnel syndrome"*[TIAB] NOT *Medline*[SB]) OR *"carpal tunnel syndrome"*[MeSH Terms] OR *carpal tunnel*[Text Word]))) OR *"Carpal Tunnel Syndrome/diagnosis"*[MeSH]) AND (*"Sensitivity and Specificity"*[MeSH] OR *"Predictive Value of Tests"*[MeSH] OR *"Comparative Study"*[MeSH])

Questions 3, 4 & 5: (*"Diagnosis"*[MeSH] OR *"diagnosis"*[Subheading] OR (*"diagnosis"*[Subheading] OR *"diagnosis"*[MeSH Terms] OR *diagnosis*[Text Word])) AND (*"Carpal Tunnel Syndrome"*[MeSH] OR *"Carpal Tunnel Syndrome"*[All Fields] OR *"Median Neuropathy"*[All Fields] OR (*"carpal tunnel syndrome"*[TIAB] NOT *Medline*[SB]) OR *"carpal tunnel syndrome"*[MeSH Terms] OR *carpal tunnel*[Text Word])) OR *"Carpal Tunnel Syndrome/diagnosis"*[MeSH]

Additionally, a list of potentially relevant studies was provided by the Work Group members. These citations were screened in the same manner as those identified by electronic searches.

b. Exclusion criteria

During Phase I screening (see [Figures 1-4](#)), all abstracts were downloaded, reviewed and evaluated for the following exclusion criteria:

- Reviews, practice guidelines, meta-analyses (except those regarding diagnosis).
- Letters, case reports, historical articles, editorials, and commentaries.
- Abstracts and unpublished study reports.
- Non prospective studies.
- Animal or *in vitro* studies.
- Cadaveric studies.
- Studies written in languages other than English.
- Studies with < 10 patients.
- Studies with patients under 21 years of age.
- Studies where gender is restricted.
- Studies where limb temperature was not monitored during electrodiagnostic tests.
- Studies where results for CTS population cannot be separated from results from other populations.
- Studies not pertaining to diagnosis of CTS.

c. Inclusion criteria

Full articles were retrieved for all abstracts passing Phase I screening. The articles then underwent Phase II screening, which consisted of evaluating the articles for the following inclusion criteria:

- Studies that meet this review’s reference standard (defined as signs and symptoms and nerve conduction study outcomes consistent with CTS) to confirm the diagnosis of CTS. (Questions 1 & 2)
- Studies addressing any diagnostic test to establish or support a diagnosis of CTS.
- The following study designs: observational [cohort, case-control, and cross sectional (XS)], or interventional [RCTs, non-randomized controlled trials (nRCTs), XS].
- Studies that compare a minimum of two diagnostic tests. (Questions 1 & 2)
- Studies where the limb temperature of the CTS patient is continuously monitored during electrodiagnostic testing according to the American Association of Electrodiagnostic Medicine Practice Parameter (Jablecki et al. 2002)
- Studies where data can be abstracted for statistical analysis.
- Studies reporting at least one of the following specific interventions:
 - Open or Endoscopic Carpal Tunnel Release. (Question 4)
 - Splinting, steroid injection or change in lifestyle. (Question 5)
- Studies reporting at least one of the following specific outcomes:
 - Post surgical improvement or resolution of CTS signs and symptoms, test results, or patient satisfaction. (Question 4)
 - Post treatment improvement or resolution of CTS signs and symptoms, test results, or patient satisfaction following splinting, steroid injection or change in lifestyle. (Question 5)

d. Grading the evidence

The quality of the evidence was graded according to five levels (see [Appendix B](#)). All eligible studies were assigned an initial level of evidence at the time of data extraction. After a more critical appraisal a final level of evidence was assigned. Studies can be downgraded according to the severity of their methodological flaws. For example, a randomized control trial that did not employ blinded assessments of test results may be re-valued as an uncontrolled clinical study. The downgrading of the formal evidence level of a study indicates the discrepancy between claims of the study authors and the results of the critical appraisal process. A limitation of this process is the potential subjectivity of the grading process; however, this was resolved, where possible, through the consensus of two independent reviewers.

e. Data extraction

Four reviewers completed data extraction independently for all studies, except for studies where data were extracted by one reviewer and checked by another. Any disagreements were resolved by consensus. Evidence tables were constructed to summarize the best evidence pertaining to each key question (see [Evidence Tables 1-21](#)).

The following data categories were extracted into electronic forms in Microsoft® Access or Excel (see [Appendix C](#)).

- Study characteristics (authors, publication year, study design, study duration, follow up period, total number of patients enrolled or hands tested, diagnostic test or treatment intervention, level of evidence).
- Patient characteristics (age – mean, median, and range, gender distribution, criteria for diagnosing CTS, duration of symptoms – mean and range, patient exclusion criteria, severity of CTS – mild, moderate or severe).
- Diagnostic tests (2 X 2 tables – number of true positives, true negatives, false positives, and false negatives, sensitivity, specificity, negative and positive likelihood ratios, negative and positive predictive value, percent of patients with positive test results, pre- and post treatment test results – mean and standard deviation)
- Treatment outcomes (type of outcomes, change in outcomes – mean and standard deviation, percent of patients with positive treatment outcomes).

f. Analysis

The purpose of the statistical analysis was to assess the diagnostic accuracy of various tests commonly used to diagnose carpal tunnel syndrome. Measures of diagnostic accuracy are based on the comparison of a test with a reference standard that determines the presence or absence of CTS. For this analysis, signs and symptoms of CTS and electrodiagnostic test (Questions 1 & 2); symptoms of CTS (Question 3); surgical outcomes from open or endoscopic carpal tunnel release (Question 4); and disease status (Question 5) were used as the “gold” standards.

In order to be considered for the analysis, studies had to report outcomes in terms of the sensitivity and specificity or had sufficient information on the performance of the test regarding the true positive and true negative outcomes (or likelihood ratios) in order to calculate sensitivity and specificity. Studies also had to have tests, outcome measures or durations of follow up in common to perform meta-analysis. Given the paucity and heterogeneity of the data for specific questions, we did not perform meta-analytic techniques in all circumstances. When possible, effect sizes were pooled across different studies, and heterogeneity was assessed with I-squared statistic (Higgins & Thompson 2002).

We attempted to meta-analytically construct receiver operating characteristic (ROC) curves for each diagnostic group or individual tests where sufficient data were available. These curves described how the test's performance in those with CTS (sensitivity or true positive rate) varies with its performance in those without CTS (false positive rate or 1 - specificity). The area under the curve (AUC) provides a measure of the overall accuracy of a test. All ROC curves were calculated using Meta-DiSc version 1.4 (Zamora et al. 2006) and Comprehensive Meta Analysis version 2 (Borenstein et al. 2005). Due to unexplained heterogeneity, we did not complete these meta-analyses.

Meta-analyses were also performed to pool the clinical outcomes of patients treated surgically with carpal tunnel release and to compare different individual and groups of diagnostic tests. Based on available data, meta-analyses were conducted for the diagnosis and surgery studies to determine whether clinical, electrodiagnostic, or clinical plus electrodiagnostic test results were associated with surgical outcomes. A meta-analysis of carpal tunnel surgery data examined pre-post surgery standardized mean differences in electrodiagnostic test results. Meta-regressions that consider diagnostic tests as predictors of surgical outcomes were examined as well. These meta-regressions employed the permutation method of Higgins & Thompson 2004. Several subgroup analyses were performed to identify factors that may be related to diagnostic variations. All meta-analyses and meta-regressions were performed using Stata 9.2 (StataCorp LP, College Station, Texas).

g. Guideline recommendations

The 8 member Work Group of experts in orthopaedic surgery, plastic surgery, physical medicine and rehabilitation, and electrodiagnostic medicine proposed recommendations for the diagnosis of carpal tunnel syndrome. AAOS staff provided identification and critical appraisal of key studies, and ratings of the quality of the evidence that correspond to each recommendation. The resulting summary of proposed conclusions and recommendations for consideration was presented and deliberated upon by members of the Work Group in a meeting on February 24, 2007.

A grading system for recommendations was employed to assist Work Group members in assessing the entire body of evidence (rather than an individual study) and indicating the strength of the recommendation. Each guideline recommendation was graded using the following scale:

- A: Good evidence (Level I Studies with consistent finding) for or against recommending intervention.
- B: Fair evidence (Level II or III Studies with consistent findings) for or against recommending intervention.
- C: Poor quality evidence (Level IV or V) for or against recommending intervention.
- I: There is insufficient or conflicting evidence not allowing a recommendation for or against intervention.

Voting on guideline recommendations was conducted using a modified nominal group technique (Murphy et al. 1999). Consensus was reached on final recommendations.

III. Results

The charge to review all clinically available methods for the diagnosis of CTS was broad. The initial search in PubMed yielded 424 citations for questions 1 and 2, and 1,710 citations for questions 3, 4 and 5. An additional 234 citations were identified from a manual search of reference lists and from studies provided by the Work Group members. After screening these citations, 92 studies were potentially eligible for data extraction (listed in [Appendix E](#)). Many of these studies reported multiple assessments (usually either several diagnostic methods or

several variations of one diagnostic method, frequently using the same sample of patients or hands). Figures 4.1-4.5 display the distribution of the assessments across the categories of signs and symptoms, clinical, and electrodiagnostic methods considered in this review.

Evidence Table 21 summarizes the number of studies excluded during Phase II screening or data extraction, organized by exclusion reason. The common reasons for exclusion included: limb temperature not reported (k=198); insufficient or not extractable data (k=123); not relevant (k=92); AAOS defined reference standard is not met (k=37); gender restricted (k=34); retrospective study design (k=32); does not compare two or more tests (k=19); insufficient control group size (k= 13); and < 10 patients total sample size (k=11). Less frequently cited reasons for exclusion (k < 10) included: review articles, case reports, electrodiagnostic test not examined, unresolved statistical problem(s), and duplicate studies.

Figures 5-8 depict the quality (level of evidence) of studies included in the review for each key question based on the type of study design. There were no Level I studies that evaluate diagnostic tests based on prospective controlled trials that measure patient outcomes in groups that have received or not received a particular test. Therefore, systematic reviews of Level I studies were non-existent. Consequently, like most other diagnostic tests, tests for diagnosis of CTS were assessed using an evidence base of data from Level II cohort, case-control and cross-sectional studies. The majority of studies, however, provided Level III or IV evidence (k=61).

When the individual diagnostic tests in cohort studies were compared against a 'gold standard' based on signs/symptoms and electrodiagnostic testing, Phalen's test had the best combination of sensitivity and specificity (80% and 91%) (Raudino et al. 2000, Evidence Table 3). In general, the sensitivity of diagnostic tests ranged from 4 to 98 percent, with specificities of 8 to 100 percent. Ranges for positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+) and negative likelihood ratio (LR-) also varied widely.

When compared with symptoms as the reference standard, the sensitivity of electrodiagnostic tests or nerve conduction studies ranged from 11 to 99 percent, with corresponding specificities of 13 to 100 percent (Evidence Table 11). This variation in test performance is due to heterogeneity among the studies. Potential sources of heterogeneity relate to the preponderance of case control study designs, spectrum of the target condition in the CTS population, and the diagnostic test methods used. Among the electrodiagnostic tests from cohort studies, median sensory nerve conduction velocity and median-ulnar latency difference had the best combination of sensitivity and specificity (94% and 100%, Padua et al. 1997; 86% and 83%, Nathan et al. 1993).

Plots in ROC space of all tests, diagnostic groups or individual tests are shown in Figures 10-94. If an imaginary 45-degree diagonal line were drawn from lower left to upper right, it would represent tests that are of no diagnostic value; such tests would detect CTS no better than chance. The top left corner represents a perfect test. The higher an ROC curve rises toward the top left corner, the more useful the test would be (both the sensitivity and specificity increase toward the top left).

We quantified how the ROC curve rises to the upper left hand corner by measuring the area under the curve (AUC); normally the larger the area, the better the diagnostic test. An area of 1.0 indicates an ideal test because it achieves both 100% sensitivity and 100% specificity. A test with an area of 0.5 has effectively 50% sensitivity and 50% specificity and indicates no discriminating ability. In practice, a diagnostic test is going to have an area somewhere between these two extremes. The closer the area is to 1.0, the better the test is, and the closer the area is to 0.5, the worse the test is. Overall, our analyses of the performance of the diagnostic signs and tests against a ‘gold standard’, defined by signs & symptoms and electrodiagnostic testing, was near or worse than chance with AUC’s ranging from 0.68 to 0.06. We realize that estimates of diagnostic accuracy of signs and tests in CTS can be largely different depending on the reference standard used to define disease; however, there was no other standard available to conduct alternate ROC analyses.

Study design varied widely among studies that sought to compare the performance of diagnostic tests with a reference standard (see [Figure 9](#)). The effects of study design on the evaluation of diagnostic accuracy of symptoms, electrodiagnostic, and clinical test in CTS are shown in the comparison of ROC curves derived from case-control, cohort, and cross sectional studies (see [Figures 11-15; 56-75; 85-94](#)). The plots in ROC space in diagnostic case-control studies, where selected samples of patients already known to have the disease are compared with separate groups of normal/healthy people known to be free of the disease, show exaggerations of both sensitivity and specificity. Due to this selection spectrum bias, diagnostic accuracy estimates obtained from case-control studies including well-defined groups of subjects with or without CTS may not be applicable to the clinically relevant population.

According to the ROC curves shown in [Figures 16-28](#), CTS is difficult to predict from the following clinical tests (when compared to a reference standard comprised of signs, symptoms of CTS and electrodiagnostic tests): clinical/psychomotor (Flick, Thenar sign, and pinch strength), provocative (carpal compression, Phalen’s test, Tinel’s sign, and pressure provocative), sensory (touch tests, vibrometry, Semmes-Weinstein monofilament, 2-point discrimination, and ridge/gap detection), and systematic. We found similar ROC curves among electrodiagnostic tests when compared to symptoms as the reference standard (see [Figures 76-84](#)). These tests generally follow a diagonal path from the lower left hand corner to the upper right hand corner. This means that every improvement in false positive rate is matched by a corresponding decline in the false negative rate, and would detect CTS no better than chance. A few tests, such as pinch strength, 2-point discrimination, ridge detection, and sensory ulnar velocity had areas under the ROC curve of less than 0.5, which represent a diagnostic performance that is worse than chance (see [Figures 18, 27, and 37](#)).

Although electrodiagnostic/nerve conduction studies are often considered the definitive diagnostic test for CTS, findings in the literature do not consistently support this view. Nine of 25 studies concluded that preoperative nerve conduction testing is not a useful indicator of surgical outcomes (Patiala 1985, Nau 1988, Braun 1994, Glowacki 1996, Padua 1997, Finsen 2001, Rege 2001, Vogt 2002, Prick 2003) (see [Figure 95](#)). Because no data was provided concerning the sensitivity, specificity, or predictive value of clinical and electrodiagnostic

studies compared with successful surgically treated CTS, we were not able to plot ROC curves for these studies. Calculation of specificity from the published results was not possible because negative results were generally not reported on surgery.

Since electrodiagnostic/nerve conduction studies are usually considered a reference standard for diagnostic tests in published studies of CTS, validation of the results of available studies is difficult. This is because validation requires an independent reference standard to which one can compare the test results. The only reference standard that would meet this criterion would be confirmation of CTS with operative findings of either open or endoscopic transaction of the transverse carpal ligament. The surgical studies that reported diagnostic results reported measures of clinical and functional status preoperatively and post-operatively (see [Evidence Table 5](#)). Correlation of the outcome measurements with clinical test results is the subject of key question 4a and correlation of these measurements with a combination of clinical and electrodiagnostic tests is the subject of question 4b. These studies demonstrated that clinical tests detected a greater number of false positives than did electrodiagnostic tests. Conversely, electrodiagnostic tests had more false positives than did clinical tests (see [Figures 96-97](#)).

Five studies used combination tests where both clinical and electrodiagnostic tests were obtained preoperatively (see [Figures 98-99](#)). Two studies reported detection rates for both tests, under the assumption that a positive result by either test was a true positive. Three studies used either a positive clinical or electrodiagnostic test to confirm abnormal findings; in these studies the test results provided the definitive diagnosis. Although CTS can be diagnosed primarily on clinical or electrodiagnostic grounds, these studies provide evidence that a combination of clinical and electrodiagnostic tests for CTS is more sensitive than the singular test, without a decrease in specificity. A meta-regression of surgical outcomes and combination tests showed significant relationship between the two ($p < 0.05$), i.e., combination tests behaved like the ‘gold’ standard – positive surgical outcomes. Also, the evidence shows that combination tests can better provide treatment orientation vis-a-vis surgical carpal tunnel release.

Although there is evidence that combination tests provide enhanced detection of CTS, there are several issues that must be addressed before it can be concluded that combination tests should replace conventional single test for routine CTS screening. With the exception of one study where use of the combination test as an adjunct to conventional testing resulted in an alteration in surgical management, none of the studies reviewed demonstrated that use of combination tests alters net surgical outcome. Moreover, these combination tests vary considerably among the studies: Haupt et al. 1993 (Motor and sensory tests) + (distal motor latency of the median nerve, antidromic sensory nerve conduction velocity, EMG examination of abductor pollicis brevis); Braun et al. 1994 (range of motion, grip strength, pinch strength, monofilament sensory evaluation, Phalen’s, Tinel’s) + (Sensory latency over 3.5 ms); Glowacki et al. 1996 (Phalen’s or Tinel’s) + (Motor latencies > 4.0 ms, sensory latencies > 3.7 ms, amplitudes $< 20\mu\text{V}$, or a conduction velocity $< 50\text{m/s}$ with evidence of fibrillation); Boniface et al. 1994 (sensory tests) + (prolonged median sensory conduction velocity, distal motor latency to abductor pollicis brevis). We have inadequate evidence, given the small number of studies in these diagnostic groupings, to recommend an optimal

combination of diagnostic tests for CTS. Since this review did not include economic assessments, it is unclear whether the use of combination tests would lead to an increase or a reduction in the cost of diagnosis.

Studies on splinting, steroid injection, and lifestyle changes reported many different interventions, durations of intervention, reference standards, and outcome measures for patient evaluation. This made it impossible to compare the findings among these studies. In addition, some studies did not use validated questionnaires on patient health outcomes, and most studies were not blinded. Collectively, studies using functional and severity outcome measures suggested that most patients improved at one month after steroid injections or splinting (see [Figures 100-101, 103](#)). Electrodiagnostic test results, however, indicated that the largest percentage of patients improved at three months (using change in distal motor latency score (ms) as an indicator of improvement) following the same treatment modalities (see [Figure 102](#)).

Selection spectrum bias is common among studies of diagnostic accuracy of MRI for CTS tests. Most studies either comprised cases that already had been diagnosed clinically with CTS, which tended to inflate estimates of sensitivity, or chose completely healthy individuals as controls, which introduced bias in the estimates of specificity. With this information at hand we tried to minimize bias in this review by using strict inclusion criteria that eliminated many poor quality studies (20 of 26 studies excluded). We included both case-control studies and cohort designs but grouped them separately (k=6). Studies were only included if there were consecutive patient recruitment, an adequate description of the test under study and the reference test were provided, and both the cases and controls had to have had the same reference test (i.e., clinical/electrodiagnostic) applied at the same definition or level.

The case-control design in 5 studies for MRI has a collection of very obviously diseased patients among the cases and very obviously healthy patients among the controls (Seyfert et al. 1994, Kliendienst et al. 1996, Keberle et al. 1999, Zagnoli et al. 1999, Jarvik et al. 2002, Musluoglu et al. 2004). The spectrum bias and the lack of patients in the ambiguous middle of the spectrum in these studies often overstated the sensitivity and specificity of MRI as a diagnostic test for CTS. We identified a tendency towards greater variability in specificity among studies, compared with sensitivity. The analyses also demonstrated significant heterogeneity, making pooled estimates impossible (see [Figures 104-106](#)). Nonetheless, estimates of the overall magnetic resonance median nerve imaging indicate that these tests have low specificity (54%), although the sensitivity was quite high but not likely much higher than 91%. These findings suggest that MRI would be inappropriate as a single test for CTS, but may be useful in combination with clinical and/or nerve conduction studies.

We located no evidence indicating a relationship between computerized axial tomography (CAT) and electrophysiological findings. Two studies recommended against the use of CAT in the preoperative assessment of patients with CTS, citing insufficient sensitivity to show changes in symptomatic patients, but the studies do not present adequate data or statistical findings in support of this observation (Merhar et al. 1986, Marshall and Davies 1990).

IV. Limitations of the evidence base

The major limitations of this systematic review are related to weaknesses inherent in the available published literature on the diagnosis of CTS. For each key question the highest level of evidence (randomized controlled trials (RCTs) and systematic reviews of RCTs) was initially sought. Due to the paucity of such studies, non-randomized comparative trials (e.g., cross-sectional, case-control, cohort studies) were also considered for all questions.

Frequently, these studies were subject to spectrum bias and there was no blinded assessment of test results. In several studies there was no statistical analysis and the follow-up periods differed between the groups. Methodological rigor of the included studies, therefore, was low to medium.

Many studies were excluded from this review due to insufficient data or unreported limb temperature during electrodiagnostic testing (see [Appendix D](#)). While strict application of inclusion and exclusion criteria caused some pertinent and potentially useful studies to be excluded, this review sought to apply uniform criteria that were established *a priori*. Even with these restrictions, a sufficient number of studies met inclusion criteria to address most of the key questions.

Another limitation of this review is that it was limited to published studies only. As studies with unfavorable results are often not published, the accuracy of a reference standard, such as positive surgical outcome, may appear falsely elevated. Test of publication bias conducted among a few studies of electrodiagnostic testing, however, did not find the presence of publication bias (Egger and Smith 1998).

Case-mix (selection or spectrum) bias occurs when cases are selected that inaccurately reflect the range of cases that occur in the general population. This was particularly true for case-control studies in which case selection generally favored selecting patients with the most advanced disease and the healthiest controls. A study using these easier cases to diagnose is more likely to show a favorable result than when the diagnostic test is used in general practice.

While most studies reported diagnostic results, these results were reported in a wide variety of formats. The variable quality and diversity of tests interfered with the ability to statistically amass a coherent body of evidence. Treatment outcomes, such as functional and surgical status, were described with such wide variation that the results from different studies could not be readily combined in a meaningful way to determine if they correlated with the diagnostic test of interest.

References

- 1) Black N, Murphy M, Lamping D, McKee M, Sanderson C, Askham J, Marteau T. Consensus development methods: a review of best practice in creating clinical guidelines. *J.Health Serv.Res.Policy* 1999 Oct;4(4):236-48.
- 2) Borenstein M, Hedges L, Higgins J, Rothstein H. *Comprehensive Meta-analysis Version 2*, Biostat, Englewood NJ (2005).
- 3) Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann.Intern.Med.* 1997 Mar 1;126(5):376-80.
- 4) Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001 Jul 21;323(7305):157-62.
- 5) Egger M, Smith GD. Bias in location and selection of studies. *BMJ* 1998 Jan 3;316(7124):61-6.
- 6) Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *CMAJ.* 1986 Mar 15;134(6):587-94.
- 7) Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat.Med.* 2002 Jun 15;21(11):1539-58.
- 8) Higgins JP, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat.Med.* 2004 Jun 15;23(11):1663-82.
- 9) Jablecki CK, Andary MT, Floeter MK, Miller RG, Quartly CA, Vennix MJ, Wilson JR. Practice parameter: Electrodiagnostic studies in carpal tunnel syndrome. Report of the American Association of Electrodiagnostic Medicine, American Academy of Neurology, and the American Academy of Physical Medicine and Rehabilitation. *Neurology* 2002 Jun 11;58(11):1589-92.
- 10) Mulrow CD, Cook DJ, Davidoff F. Systematic reviews: critical links in the great chain of evidence. *Ann.Intern.Med.* 1997 Mar 1;126(5):389-91.
- 11) Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat.Med.* 2002 Jun 15;21(11):1559-73.
- 12) Zamora J, Abaira V, Muriel A, Khan KS, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Medical Research Methodology* 2006, 6:31.