# Chapter 2. Methodology

## Conditions of Interest

This evidence report is concerned with worker-related upper extremity disorders. The term "worker-related" implies a disorder that affects workers, not a disorder caused by work. In this report, we address four specific disorders: (1) carpal tunnel syndrome, (2) cubital tunnel syndrome, (3) epicondylitis, and (4) de Quervain's disease. This list of disorders was determined during discussions among ECRI, the Agency for Healthcare Research and Quality (AHRQ), the organizations that nominated this topic to AHRQ, and a panel of technical experts. Below, we provide further details about the nominating organizations and technical experts.

## Technical Experts

Technical Experts were employed to assist in defining the scope of this evidence report, developing its questions, and developing the criteria for retrieving and including articles. Seven organizations were solicited to nominate individuals who could serve as Technical Experts. All solicitations were pre-approved by AHRQ. All seven organizations nominated an individual. Thus, the Expert Panel was comprised of individuals from the American Association of Electrodiagnostic Medicine, the American Academy of Neurology, the American Academy of Physical Medicine and Rehabilitation, the American Physical Therapy Association, the Association for Repetitive Motion Syndromes, the American Association of Neurological Surgeons, and the American Academy of Orthopedic Surgeons. The participation of these individuals and organizations does not imply their endorsement of the findings of this evidence report.

## Key Questions

To determine the specific questions that this evidence report would address, a multidisciplinary team was assembled. This team included ECRI research staff, AHRQ project staff, representatives from the organizations that nominated this topic to AHRQ (the Social Security Administration and the American College of Occupational and Environmental Medicine), and the Technical Experts. The key questions for this report were decided during three conference telephone calls between ECRI, AHRQ, the experts, and the nominating organizations, as well as subsequent discussions between ECRI, AHRQ, and the nomination organizations.

The final set of key questions is comprised of 13 questions, 11 of which are separately addressed for the four above-mentioned disorders. The remaining two questions are not disorder specific. This evidence report is correspondingly organized. Thus, we first

address each of the 11 questions for each disorder, beginning with carpal tunnel syndrome, and conclude by addressing the two questions that are not disorder-specific.

## Condition-Specific Questions

The 11 condition specific questions that we address in this report are:

**Question #1:** What are the most effective methods and approaches for the early identification and diagnosis of worker-related musculoskeletal disorders of the upper extremity?

**Question #2**: What are the specific indications for surgery for worker-related musculoskeletal disorders of the upper extremity?

**Question #3:** What are the relative benefits and harms of various surgical and nonsurgical interventions for persons with worker-related musculoskeletal disorders of the upper extremity?

**Question #4:** Is there a relationship between specific clinical findings and specific treatment outcomes among patients with worker-related musculoskeletal disorders of the upper extremity?

**Question #5:** Is there a relationship between duration of symptoms and specific treatment outcomes among patients with worker-related musculoskeletal disorders of the upper extremity?

**Question #6:** Is there a relationship between factors such as patients' age, gender, socioeconomic status and/or racial or ethnic grouping and specific treatment outcomes among patients with worker-related musculoskeletal disorders of the upper extremity?

**Question #7:** What are the surgical and nonsurgical costs or charges for treatment of worker-related musculoskeletal disorders of the upper extremity?

**Question #8:** For persons who have had surgery for worker-related musculoskeletal disorders of the upper extremity, what are the most effective methods for preventing the recurrence of symptoms, and how does this vary depending on subject characteristics or other underlying health problems?

**Question #9:** What instruments, if any, can accurately assess functional limitations in an individual with a worker-related disorder of the upper extremity?

**Question #10:** What are the functional limitations for an individual with a worker-related musculoskeletal disorder of the upper extremity before treatment?

**Question #11:** What are the functional limitations of an individual with a worker-related musculoskeletal disorder of the upper extremity after treatment?

## Non-Condition-Specific Questions

The two questions that are not condition specific are:

**Question #12:** What are the cumulative effects on functional abilities among individuals with more than one worker-related musculoskeletal disorder of the upper extremity in the same limb?

**Question #13:** What level of function can one achieve in what period of time when one is required to change hand dominance as a result of injury to his or her dominant hand?

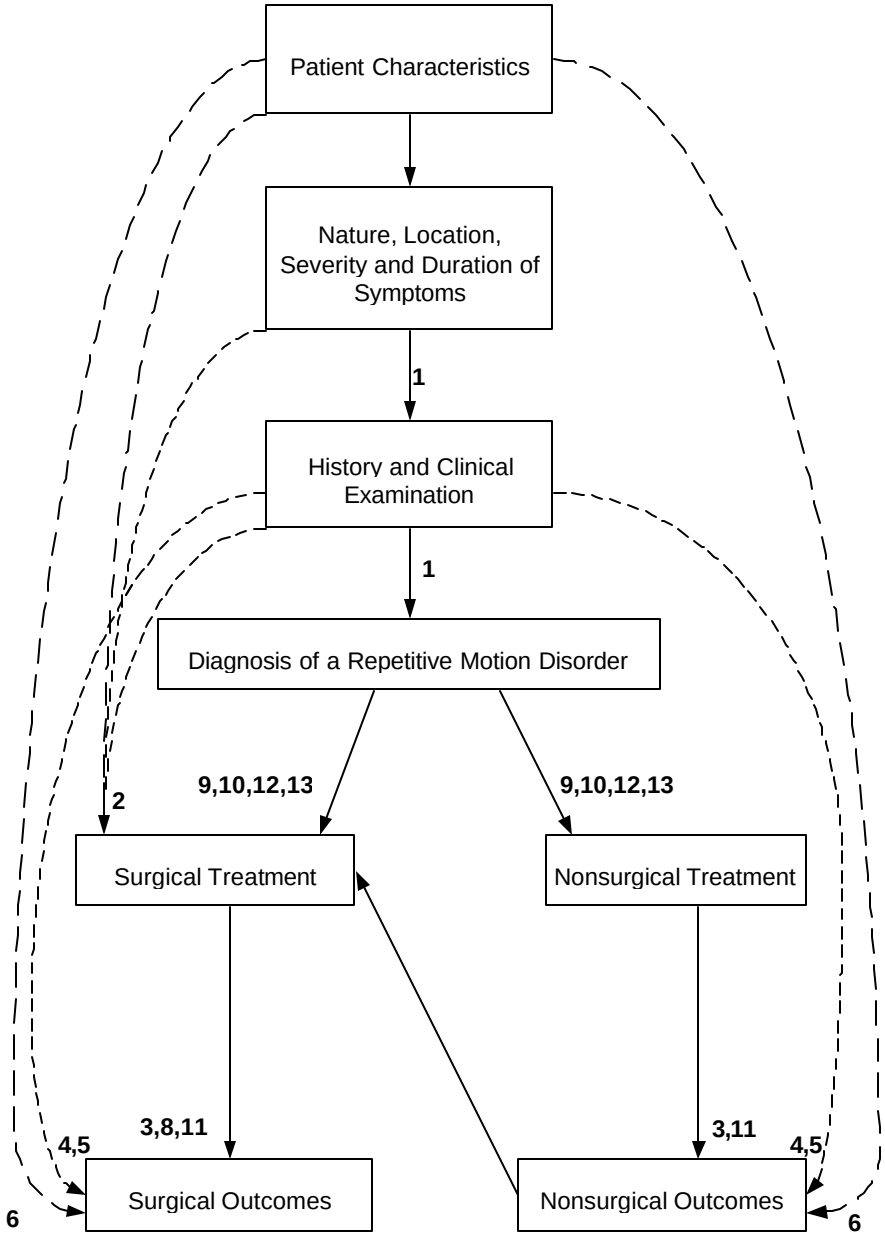
# Causal Pathway

The scope of this report can be illustrated by a causal pathway. More specifically, this pathway illustrates the key questions and the relationships among them. It also illustrates items that are beyond the scope of this evidence report. This pathway is shown in Figure 6. The rectangles in this figure depict the primary clinical "events", from presentation of a patient (who has certain characteristics that may be at least partly diagnostic and/or prognostic) to the outcomes that the patient experiences (e.g., improves/does not improve). That this, in fact, is a pathway that proceeds in a certain chronological order is depicted by solid arrows that connect the rectangles in Figure 6. Because these arrows connect two rectangles, they are termed "links." The numbers next to each link represent the numbers of the Key Questions that address that link. Key Question 7 is not shown in the pathway because it is concerned with costs and, therefore, is not part of the clinical pathway.

The dashed lines in the figure "overarch" several rectangles. We have drawn these lines as dashed because they do not depict the sequence of events in the clinical pathway. In general, these lines portray Key Questions about how patient characteristics (including clinical findings) may influence a patient's movement through the clinical pathway or whether these characteristics influence outcomes.

Theoretically, one can derive a key question by drawing a line between any two rectangles in Figure 6. Therefore, rectangles not connected by solid or dashed lines are beyond the scope of this evidence report.

**Figure 6. Causal Pathway**




# Literature Searches

Our searches for information were designed to produce a comprehensive dataset. Therefore, we searched a number of electronic databases and other sources. These are described below.

## Electronic Database Searches

We searched 31 electronic databases. These databases were:

CISILO Database (International Occupational Safety and Health Information Centre) (through November 2000)

The Cochrane Database of Systematic Reviews (through 2000, Issue 4)

The Cochrane Registry of Clinical Trials (through 2000, Issue 4)

The Cochrane Review Methodology Database (through 2000, Issue 4)

CRISP (Computer Retrieval of Information on Scientific Projects) (through November 16, 2000)

Cumulative Index to Nursing and Allied Health (CINAHL) (1988 through September 29, 2000)

Current Contents (through December 2000)

The Database of Reviews of Effectiveness (Cochrane Library) (through 2000, Issue 4)

DIRLINE (through September 27, 2000)

ECRI Health Devices Alerts (1977 through January 2001)

ECRI Health Devices Sourcebase (through January 2001)

ECRI Healthcare Standards (1975 through January 2001)

ECRI International Health Technology Assessment (IHTA) (1990 through January 2001)

ECRI Library Catalog (through January 2001)

ECRI TARGET (ECRI's database of emerging technologies; through January 2001)

Embase (Excerpta Medica) (1974 through December 12, 2000)

ERIC (Educational Resources Information Center) (searched June 28, 2000)

Health and Psychosocial Instruments (HAPI) (through January 30, 2001)

Health Services Research Projects (HSRPROJ) (through September 27, 2000)

HealthSTAR (Health Services, Technology, Administration, and Research) (1990 through September 26, 2000)

LocatorPlus (through January 2001)

NIOSHTIC (through November 3, 2000)

Old Medline (1957 -1965) (searched September 27, 2000)

PsycINFO (1967 through January 22,2001)

PubMed (1966 through January 22, 2001)

Rehabdata (through November 2000)

SciSearch (through November 13, 2000)

U.K. National Health Service (NHS) Economic Evaluation Database (NHS EED) (through January 2001)

U.S. Health Care Financing Administration (HCFA) (through January 2001)

U.S. National Guidelines Clearinghouse (NGC) (through January 2001)

U.S. National Institutes of Health Web site (NIH) (through January 2001)

## World Wide Web Searches

To further ensure that this evidence report was comprehensive, we also searched the World Wide Web using various resources and search engines including AltaVista, NorthernLight, and Google. These resources included:

American Academy of Orthopedic Surgeons http://www3.aaos.org

American College of Occupational and Environmental Medicine (ACOEM) http://www.acoemwebapps.org/gov/welcomeNS.asp

Association for Repetitive Motion Syndromes (ARMS) http://www.certifiedpst.com/arms/

Canadian Centre for Occupational Health and Safety (CCOHS) http://www.ccohs.ca/

Centre for Clinical Effectiveness http://www.med.monash.edu.au/publichealth/cce/

Development Evaluation Committee http://www.hta.nhsweb.nhs.uk/rapidhta/main.htm

ErgoWeb http://www.ergoweb.com/

HCUPnet http://www.ahcpr.gov/data/hcup/hcupnet.htm

Medscape http://www.Medscape.com

National Institute for Occupational Safety and Health (NIOSH)
http://www.cdc.gov/niosh/homepage.html

NHS Centre for Reviews and Dissemination http://www.york.ac.uk/inst/crd/welcome.htm

Safety and Health Statistics, Bureau of Labor Statistics http://stats.bls.gov/oshhome.htm

SUM Search http://sumsearch.uthscsa.edu/searchform4.htm

TRIP Database http://www.tripdatabase.com/

## Other Sources

In addition to the above searches, we also reviewed the bibliographies and reference lists of all studies included in this evidence report, searched Current Contents—Clinical Medicine on a weekly basis, and routinely reviewed over 1,600 journals and supplements maintained in ECRI's collections.

## United States Cost/Reimbursement Data

We searched four additional U.S. government datasets solely to obtain information about costs. These were:

*2001 Physician Fee Schedule.* This Health Care Financing Administration (HCFA) dataset contains fees and limiting charges for physician services under Medicare in 2001.

*Median Costs for Hospital Outpatient Services Dataset.* This HCFA dataset contains median costs, by HCPCS codes, for services reimbursed under the hospital outpatient prospective payment system. The data are calculated based on 1996 hospital outpatient claims.

*Medicare Provider Analysis and Review (MEDPAR).* This HCFA dataset contains information for 100% of Medicare beneficiaries using hospital inpatient services. The data are provided by state and then by diagnostic related group (DRG) for all short stay and inpatient hospitals for fiscal years 1990-1996. Data include total charges, covered charges, Medicare reimbursement, total days, number of discharges, and average total days.

*Hospital Outpatient Prospective Payment System*. This HCFA dataset contains rules for payment of outpatient services provided by hospitals or affiliated organizations under hospital control. The system is based on ambulatory payment classifications (APCs). This classification system groups services both clinically and by resource utilization.

## Search Strategies

The systematic nature of the searches for information for an evidence report is a means of diminishing reviewer bias during the preparation of a report. This systematic nature is reflected in our strategies for searching PubMed/Medline and HCUPnet for ICD-9 procedure codes and CPT codes, diagnostic related groupings (DRGs), ambulatory related groupings (ARGs), and HCPS codes. These strategies are detailed in Appendix A.

## Article Retrieval Criteria

To be included in this evidence report, an article had to meet two sequentially applied sets of *a priori* criteria. The first set determined whether a full article would be retrieved. The second set, which was based on major study design flaws and certain elements specific to each question, determined whether a retrieved article would be included in the report. To facilitate comprehensive article retrieval, the retrieval criteria were designed to be broad.

The abstracts of articles identified by our searches were reviewed against the retrieval criteria to determine whether we would retrieve an article identified by our searches. This task was independently performed by six research analysts, each of whom individually worked on different questions. We retrieved an article whenever there was uncertainty about whether it met the retrieval criteria. We also retrieved articles when an abstract was not present in the search results, but when the title of the article suggested that it was relevant.

The criteria for article retrieval were:

- All patients, or a separately reported subset of patients in any given article, had to be diagnosed with a worker-related disorder of the upper extremity. No restrictions were placed on the patient populations in clinical trials of conservative or surgical treatments that were retrieved for this analysis. For studies addressing condition-specific key questions, patients had to be diagnosed with the specific disorder of interest.

- All controlled trials were retrieved, regardless of whether they were described as randomized or prospective. There was no cutoff date for year of publication. Included in the retrieved articles were those that compared a treatment to a placebo, sham, or untreated group and those that compared two or more treatments.

- Case series and other reports were evaluated only if published in 1980 or later. This was an arbitrary cut-off date set to exclude case series using obsolete techniques and outdated patient selection criteria.

- Case series had to enroll 10 or more patients. Studies with less than 10 patients are unlikely to be representative of the range of patients with the disorder being evaluated.

- Only English-language articles were retrieved.

## Inclusion Criteria

Once an article was retrieved, it was examined to determine whether it suffered from a major design flaw and whether it met certain question-specific criteria. When an article was excluded, the research analysts entered a unique article identifier and the reason(s) for exclusion into an electronic data abstraction form (DAF).

When an article was included, the unique identifier and details about the studies results, design, and enrolled patient population were entered in these forms. Additional details about the DAFs are provided below.

Many of our exclusions were made because an article contained a significant design flaw. To avoid redundancy, we do not list these flaws here. Rather, we provide a listing of the major design flaws used to exclude articles in the sections of this report in which we evaluate the quality of the literature. Below, we provide the inclusion criteria that are unique to each question:

**Question 1. What are the most effective methods and approaches for the early identification and diagnosis of worker-related musculoskeletal disorders of the upper extremity?**

Studies meeting the retrieval criteria were included:

- Only if they reported sensitivity and specificity or provided sufficient data to allow us to compute these measures of test performance.

- If they did not use obsolete tests (e.g., first- and second-generation CT scanners).

- Regardless of whether they were prospective or retrospective.

- Regardless of whether they contained a concurrent control group. Use of controlled and particularly randomized controlled studies is exceedingly rare in the evaluation of any diagnostic test. Often, such controls are not needed because the patients can validly serve as their own controls.

**Question 2. What are the specific indications for surgery for worker-related musculoskeletal disorders of the upper extremity?**

To address this question, we tabulated the characteristics of patients enrolled in clinical studies. Doing so does not require any particular study design, and this is reflected in our inclusion criteria. Thus, among the studies that met the retrieval criteria, we included:

- Controlled trials and case series of surgical patients

- Studies in which not all patients received surgery were included, but only if characteristics of patients receiving surgery were reported on separately.

- Studies that did not exclusively enroll patients with co-morbidities not routinely encountered during routine clinical practice (e.g., patients with amyloidosis).

**Question 3. What are the relative benefits and harms of various surgical and nonsurgical interventions for persons with worker-related musculoskeletal disorders of the upper extremity?**

Among studies meeting the retrieval criteria we included:

- Controlled studies, regardless of whether they were randomized or blinded.

- Studies that were not exclusively dedicated to comparing highly similar treatment variations (such as incision shape).

- Studies that reported on at least one of the seven key outcomes addressed in this assessment. The outcomes are: pain, function, quality of life, ability to return to work, ability to return to activities of daily living, harms, and global outcome.

**Question 4. Is there a relationship between specific clinical findings and specific treatment outcomes among patients with worker-related musculoskeletal disorders of the upper extremity?**

We evaluated controlled trials and case series that attempted to correlate patient-oriented outcomes with specific clinical findings, patient characteristics or duration of symptoms. It is not feasible to conduct randomized controlled trials that address this question because, by definition, one cannot fully randomize patients with different pretreatment clinical findings into different groups. Therefore, the inclusion criteria adopted for this question were:

- Studies that evaluated the relationship of pretreatment clinical findings and outcomes using multiple linear or logistic regression.

- Studies that statistically compared the outcomes of patients stratified across some pretreatment clinical finding.

- Studies reporting patient-level data were included when the data were presented in enough detail to allow us to perform independent multiple regression analyses.

- Studies that reported on at least one of the seven key outcomes addressed in this assessment. The outcomes are: pain, function, quality of life, ability to return to work, ability to return to activities of daily living, harms, and global outcome.

- Studies that examined a simple correlation between a given pretreatment variable and outcomes were included, even if they did not attempt to control for the effects of other predictor variables. However, we only included such studies if there were at least three studies that attempted to correlate the same outcome with the same predictor variable. We adopted the arbitrary criterion of requiring three correlational studies because, when taken individually, interpretation of such studies is difficult. This is because they do not contain information about potential inter-variable multicolinearity.

**Question 5. Is there a relationship between duration of symptoms and specific treatment outcomes among patients with worker-related musculoskeletal disorders of the upper extremity?**

The criteria used for this question were identical to those used for Question 4.

**Question 6. Is there a relationship between factors such as patients' age, gender, socioeconomic status and/or racial or ethnic grouping and specific treatment outcomes among patients with worker-related musculoskeletal disorders of the upper extremity?**

The criteria used for this question were identical to those used for Question 4.

**Question 7. What are the surgical and nonsurgical costs or charges for treatment of worker-related musculoskeletal disorders of the upper extremity?**

Cost and charge information from large national databases was included.

**Question 8. For persons who have had surgery for worker-related musculoskeletal disorders of the upper extremity, what are the most effective methods for preventing the recurrence of symptoms, and how does this vary depending on subject characteristics or other underlying health problems?**

- Controlled trials of any design (RCTs, prospective non-randomized, and retrospective) were included.

**Question 9. What instruments, if any, can accurately assess functional limitations in an individual with a worker-related disorder of the upper extremity?**

For inclusion in this question, a study meeting the retrieval criteria had to be:

- A case series or controlled study that measured the validity, response to treatment, or test-test reliability of the assessment instrument.

- A study not exclusively devoted to measuring the internal consistency of an instrument. Although internal consistency is important in instrument development, it does not directly address the ability of an instrument to predict functional limitations.[89]

- A study of an instrument designed to evaluate patient function. Instruments that only evaluated symptoms or that were primarily designed to aid in diagnosis were not included.

- A study of an instrument that enrolled patients with one of the four specific disorders of interest.

**Question 10. What are the functional limitations for an individual with a worker-related musculoskeletal disorder of the upper extremity before treatment?**

In addressing this question, we tabulate functional limitations. Answering this question does not require randomized controlled trials. Therefore, our inclusion criteria for studies meeting the retrieval criteria were:

- All studies, regardless of design

- Studies that measured functional disability using one of the instruments identified in Question 9

- Studies that exclusively enrolled patients with one of four conditions of interest.

- Studies reporting on functional ability using portions of these instruments or minor variations of these instruments were included as well.

- Study must not have enrolled patients who received prior treatment.

**Question 11. What are the functional limitations of an individual with a worker-related musculoskeletal disorder of the upper extremity after treatment?**

This question is similar to Question 10 and, therefore, identical inclusion criteria were employed except for the one requiring that patients must not have had prior treatment. To be included for Question 11, the study must have been of patients who received prior treatment.

**Question 12. What are the cumulative effects on functional abilities among individuals with more than one worker-related musculoskeletal disorder of the upper extremity in the same limb?**

The criteria for this question were identical to those used for Question 11, except that the study must have reported data on the patient population relevant to Question 12.

**Question 13. What level of function can one achieve in what period of time when one is required to change hand dominance as a result of injury to his or her dominant hand?**

This question also does not depend on randomized controlled trials. Therefore, we included any retrieved study, regardless of design, that employed any test of functional

ability in patients required to change hand dominance as a result of injury to the dominant hand.

## Electronic Data Abstraction Forms

Data from all articles that met our inclusion criteria were abstracted using electronic data abstraction forms. These forms were created using Microsoft Access. Using this software, separate data abstraction forms were designed for entering data about basic trial design information; patient signs, symptoms, comorbidities, characteristics, and treatments; reporting of treatment outcomes; surgical complications; and nerve conduction measurements. The data abstraction forms are presented in the appendix B.

The abstraction form for trial information contained information on trial design, purpose, author, year of publication, general diagnosis of patient condition, a specific description of the treatment outcomes examined, inclusion and exclusion criteria, and other important information with which to judge the quality of the trial. One record containing a unique trial identification number appears for each trial entered in the database.

The abstraction form for patient characteristics and treatments was designed to contain information on each patient group within a trial. A separate record containing a unique patient group identification number appears for each patient group within a trial. This form contained entries for treatment given to the patient group, stratification of patient groups based on pretreatment characteristics, number of patients in the group, specific descriptions of patient treatment, and patient characteristics such as age, dropouts, signs, symptoms, disease severity and duration of symptoms prior to treatment.

Abstraction forms with similar design were created to contain information on treatment outcomes. Separate abstraction forms were needed for dichotomous, categorical and continuous outcome data. These forms contained entries for the patient group identification number, number of patients reporting the outcomes, and time the outcome was measured. A separate record was entered for each patient group and each follow up time for which an outcome was reported.

Special forms were designed for symptoms, comorbidities, complications, and results of diagnostic tests.

Because diagnostic trials differ from treatment trials in many important ways, several special forms were used in the abstraction of diagnostic data, and irrelevant sections of the other data abstraction forms were not completed.

One clinical trial information form and one diagnostic clinical trial information form were completed for each study; not all of the fields in the clinical trial information form were relevant to the diagnostic studies. One patient groups—diagnostics and characteristics form was completed for each patient group or subgroup in each study. Most articles from which we abstracted data reported on two groups; some reported more. One diagnostic test information form was completed for each diagnostic test result

reported in each study. Because separate forms were completed for each test parameter reported (e.g. distal motor latency v. distal sensory latency), most studies required more than one form and several required 30 or more forms. One study reported 57 different tests.[90]

## Articles Identified

Our searches identified 7,312 articles. Of these, 1270 were clinical trials. The number of articles included for each question is shown in Table 2.

**Table 2. Number of articles Included for Each Key Question**

| Question # | Carpal Tunnel | Cubital Tunnel | Epicondylitis | De Quervain's |
|---|---|---|---|---|
| 1 | 189 | 20 | 10 | 0 |
| 2 | 145 | 32 | 19 | 3 |
| 3 | 44 | 3 | 50 | 1 |
| 4 | 12 | 11 | 3 | 1 |
| 5 | 5 | 14 | 7 | 1 |
| 6 | 21 | 15 | 6 | 1 |
| 8 | 0 | 0 | 0 | 0 |
| 9 | 8 | 0 | 3 | 0 |
| 10 | 2 | 0 | 2 | 0 |
| 11 | 12 | 0 | 0 | 0 |

For the two questions that were not condition specific, Questions 12 and 13, we included 0 and 2 articles, respectively. Question 7 is not depicted in the above table because we addressed it using information from a national database, not published articles.

## Evaluating Literature Quality

Because this is a "best evidence" synthesis, we incorporated studies that represented the best available evidence, not the best possible evidence. Therefore, not all evidence that we included is of equal quality.

The quality of studies of treatments that we evaluated can be ranked according to the following hierarchy:

Randomized controlled trials

Other prospective controlled trials

Retrospective controlled trials, including those with historical control groups

Prospective case series

Retrospective case series

The hierarchy, like any evidence hierarchy, is only a rough guide. As noted above, randomized controlled trials are not necessary for some of the questions (among which are questions about diagnostics) that we addressed. In such cases, this hierarchy is not applicable. Therefore, for these questions, we discuss the dimensions along which we evaluated the quality of the literature when we address that question. These discussions appear in the appropriate Internal Validity sections under each of these questions.

# Statistical Methods

## Meta-Analysis of Studies of Treatment

Meta-analyses of studies of treatments were conducted using Hedges' d as a measure of each study's effect size, and then computing the precision-weighted summary d from the combined results of all studies.[91] Hedges' d is the difference between the means of any study's two groups expressed in standard deviation units. We performed meta-analyses on data from studies of treatments only when four or more controlled studies of a given treatment reported the same outcome. We did not perform meta-analyses of smaller data sets because of the high potential for publication bias to affect their results.

For computation of effect sizes derived from dichotomous outcomes, we converted the odds ratio to Hedges' d as described by Hasselblad and Hedges.[92] For computation of effect sizes derived from rating scale data, we calculated a mean for each group as described by Torgenson (his equations 71-78).[93] An advantage of this method is that it does not assume that all patients employ exactly the same boundaries for each category in a rating scale.

We employed two tests for heterogeneity, the Q statistic and each study's standardized residual. We regarded the data as heterogeneous if the results of either test was statistically significant. When we detected heterogeneity, we analyzed the data for sources of heterogeneity. It was not always possible to find a source, particularly when there were only a small number of studies in the meta-analyses. These models were computed using a modified method of moments.[94] To further assist in interpreting the results of our meta-analyses, we present the results of our fixed effects models in terms of Forrest plots and as a pair of normal curves. Each curves represents the distribution of results in a study's two groups. The difference between the means of these two normal curves represents d, the effect size. We quantified the degree of the non-overlap of these two curves using the $\cup$ statistics described by Cohen[95], and have expressed these results in terms of the overlap between these curves.

## Meta-Analysis of Diagnostic Studies

Diagnostic test meta-analyses were done according to the method of Littenberg and Moses.[1] Meta-analyses of diagnostic studies were performed only when there were 10 or more retrieved trials of a given test. We adopted this criterion to ensure that this

evidence report would focus on the tests for which there is the greatest research interest. We have taken the mean threshold as the best estimate of a single threshold, and the values of sensitivity and specificity at the mean threshold as the single best global estimate of test effectiveness.

Before using the results of a meta-analysis, we verified that there was no statistically significant heterogeneity among the results of the included articles. This was accomplished, using the Q statistic, as described by Hasselblad and Hedges.[92] The presence of heterogeneity indicates that something other than threshold is affecting sensitivity and specificity, and that the points on an ROC curve are not derived from the same population of sensitivity/specificity pairs. If heterogeneity was detected, we removed any subgroups that caused the heterogeneity from the analysis. If there were no subgroups in the analysis, or those subgroups did not cause the heterogeneity, we looked for data points that were outliers, and reported the meta-analytic results with and without exclusion of these outliers.

Meta-analysis results of diagnostic tests are reported both in table and graphical form. Tables list each study in the meta-analysis, its 2 x 2 data, and any special steps ECRI had to take in abstracting that data. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are also reported in those tables, along with confidence intervals on each of these ratios, calculated according to Wilson's method.[96] Finally, the prevalence of WRUED cases in each study is reported. The last row of the results table provides the sensitivity/specificity at mean threshold results of the meta-analysis, along with the sensitivity and specificity of the points representing the 95% confidence interval on the mean threshold point. Summary values for PPV and NPV are not calculated in the meta-analysis because they are dependent on disease prevalence. Meta-analysis results graphs include the summary ROC itself, the confidence interval and the sensitivity/specificity data points for each included article. The diagonal line in each graph represents the performance of a test that worked no better than chance.

Some investigators based their diagnostic thresholds on results obtained in a control population of individuals without the condition, typically setting a threshold at 2.0 or 2.5 standard deviations from the mean test score of the controls. When the actual number of positive and negative results in the control subjects was reported in the article, we used that data in the meta-analysis. In cases where these numbers were not reported, we assumed a normal distribution of test results in the control subjects, and calculated the theoretical number of false positives and true negatives based on the one-tailed normal distribution. If the threshold was two standard deviations from the mean, one expects false positive results in 2.275% of controls; if the threshold was 2.5 standard deviations, then false positives should make up 0.621% of the control group. The appropriate percentage was multiplied by the number of control subjects and rounded to the closest whole number of patients to get counts for the 2 x 2 table. If the number of controls given the study test was not reported, the article was excluded from analysis even though we knew test specificity from the reported threshold. This is because actual counts of false positives and true negatives are needed to obtain confidence intervals on specificity and the predictive values.

## Other Computations

We performed numerous other statistical computations in addition to those involved in performing meta-analyses. We describe these calculations and the logic behind them in our considerations of the appropriate questions. Briefly, these calculations included:

1. Corrections for attrition; Following all patients for the duration of a study is difficult, particularly when the study is relatively long term. It is possible that in some studies, poor outcomes among patients lost to followup could overturn the results of a study, including those of a well-designed randomized controlled trial. Therefore, wherever possible, we made conservative assumptions about outcomes in patients who were not accounted for in an effort to determine how robust reported results were. This approach is preferable to one that ignores attrition and to one that discards such studies that exceed an arbitrary attrition level. The former approach could lead to incorrect conclusions and the latter can lead to information loss.

2. Statistical power analyses; Studies that do not contain a sufficient number of patients cannot detect statistically significant differences between groups, even when these differences are clinically meaningful. Therefore, whenever possible, we computed the minimum between-group difference that any given controlled study had the power to detect.

3. Multiple regression; For certain questions, the results of multiple regressions were of interest, but such analyses were not conducted by the authors. We therefore conducted these analyses when t-patient-level data were available.

4. Computations of effect sizes for all studies, when possible, even when no meta-analysis was performed. Results of statistical tests (p-values) do not convey information about the magnitude of an effect. To provide an idea about this magnitude, we computed effect sizes for all controlled studies, wherever such computations were possible.

5. Determinations of whether there were statistically significant differences between the characteristics of patients in any given study. Although studies may report that they were randomized, it is sometimes the case that the randomization protocol was not adequately followed or the study was not truly randomized. These departures from randomization can manifest themselves in pretreatment between-group differences in patient characteristics.

6. Computation of pretreatment effect sizes. Departures from randomization can also manifest themselves as a statistically significant difference in the outcome between groups prior to the administration of treatment. For example, if the pain levels experienced by patients were significantly different before treatment, one might suspect that the study was not truly randomized.

7. Verification of 2 x 2 tables reported in studies of diagnostic tests. Because peer-reviewed published articles often contain errors in reported results, we attempted to verify the calculations in each article. If an error was found, we corrected the data and included it in the analysis. If we could not verify the 2 x 2 table, the article was excluded. These exclusions are documented in the text of this report.

## Table 3. Coding of Patient Inclusion Criteria

| Code | Definition |
|---|---|
| **WRUED groups** | |
| Symptoms/presented | Patients had unspecified symptoms of the disorder being studied, or were referred for diagnosis of suspected WRUED |
| Simple signs/symptoms | Patients included if they had specified symptoms of the disorder, but other tests such as nerve conduction tests were not used for patient selection |
| Simple NCS | Patients included if they had abnormal results in a specific nerve conduction test or tests (no more than three tests in selection algorithm) |
| Complex objective standard | A specified algorithm with more than three nerve conduction studies or combining specific NCS tests with specific symptoms |
| Unspecified (diagnosed) | Authors reported that all patients had been diagnosed with the disorder in question, but did not detail how the diagnosis was defined |
| Other | Details reported in separate database field |
| **Control groups** | |
| Healthy volunteers | Subjects drawn from hospital or community populations, and not being evaluated for other upper extremity disorders |
| Workers at risk | Asymptomatic individuals considered to be at risk for WRUED |
| Unrelated disease | Subjects were being evaluated or treated for known abnormalities of the hand or wrist unrelated to WRUEDs |
| Contralateral arm | Unaffected contralateral extremity of persons with diagnosed WRUED |
| Other | Details reported in separate database field |

## Table 4. Coding of Diagnostic Test Groups

| Test group | Included tests |
|---|---|
| Imaging tests | Radiography (film x-ray), computed tomography, MRI, ultrasound |
| Nerve conduction | Amplitude, latency, and velocity of signal conduction in median and ulnar nerves |
| Composite nerve conduction | Differences and ratios of nerve conduction test results |
| Signs and symptoms | Phalen's maneuver, reverse Phalen maneuver, Tinel's sign, Durkin (carpal compression) test, sensory diagrams |
| Sensory tests | Semmes-Weinstein monofilament test, vibrometry, current perception threshold |

### Table 5. Coding of Results Reporting Level

| Reporting level | Definition |
|---|---|
| Patient-level | Results for each patient reported individually. This includes studies where patient-level results were reported in a graph rather than a table. Where possible, ECRI research analysts |
| Counts | Sufficient data to yield a two-by-two truth table relating test results to another condition (usually patient's assignment to disease or control group) |
| Summary statistics | Mean and standard deviation of results for all patients in the group |
| Agreement or difference | Statistics reporting agreement or difference between results of one test and another, but not the results themselves |
| Technical criteria | Accuracy, precision, and reproducibility of the test results, but not the results themselves. |

### Table 6. Coding of Studies of Special Interest

| Characteristic | Definition |
|---|---|
| Longitudinal data | Study reported repeated measurements on the same subjects, from which information on the progression of the condition can possibly be derived |
| Early diagnosis | Study reported that it was intended to identify early-stage disease. For purposes of this assessment, we relied on the authors' own definitions of "early diagnosis" and did not try to validate that validate that description. |
| Screening study | Study included at least one group of subjects that can be considered a screening population (e.g. asymptomatic individuals whose work entails repetitive movements). |

## Peer Review

To select peer-reviewers for the draft evidence report, ECRI prepared a list of 30 potential reviewers. This list was submitted to AHRQ, which approved all reviewers. Letters inviting these individuals to review were then mailed. Fifteen individuals responded to these letters, 12 individuals agreed to review the draft evidence report, and 9 individuals returned reviews.

Upon receipt of reviews, ECRI revised the draft report accordingly. ECRI also prepared a document describing the disposition of all substantive reviewer comments and supplied this document to AHRQ for review and approval.