

Standardized criteria for the diagnosis of carpal
tunnel syndrome: A synopsis of the
development of the CTS-6

Brent Graham, MD, FRCSC

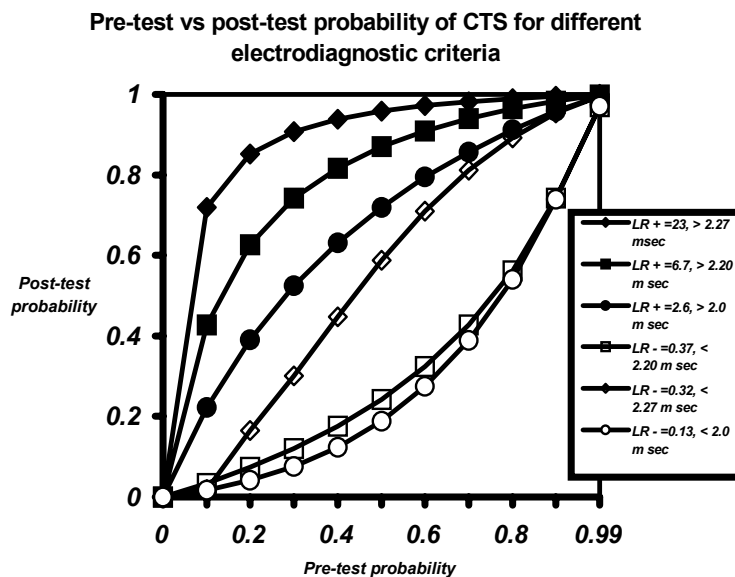
Supported by Physicians' Service Foundation Grant 97-52

The main objective of our work over the past five years has been to develop standardized criteria for the diagnosis of carpal tunnel syndrome. These criteria have been identified and comprise a diagnostic instrument that we are naming the **CTS-6**. Initial validation of this instrument has been completed. Reliability studies are currently under design. The goal of this communication is to provide participants in the study with a pre-publication synopsis of its findings.

Overview

Our clinical experience indicated that unsatisfactory results after treatment for carpal tunnel syndrome could frequently be linked to an inaccurate initial diagnosis. A review of the literature indicated that the diagnosis of carpal tunnel syndrome is often made according to highly variable criteria. A large number of internal medicine, surgical and primary care specialties treat the condition and it seemed possible that some of the variation might be specialty-specific. In other words, the way in which a rheumatologist, an occupational health physician and a neurosurgeon diagnose carpal tunnel syndrome might vary substantially. In addition, the use of electrodiagnostic tests in the overall evaluation of carpal tunnel also appeared to be highly variable. While some clinical groups reported little or no role for electrodiagnostic tests, others considered this investigation to be the gold standard for the diagnosis of carpal tunnel syndrome.

We reasoned that the role of electrodiagnostic testing should be in adjusting the probability of carpal tunnel syndrome established clinically. In a bayesian context, the clinical evaluation of carpal tunnel syndrome should allow the pre-test probability of carpal tunnel syndrome to be established. Where this is intermediate, that is, neither very high or very low, an electrodiagnostic evaluation is indicated and should be used to modify the probability of carpal tunnel syndrome either up or down depending on the outcome of testing. We tested this hypothesis on data published in the literature. The results are shown in Figure.



In the absence of widely accepted, standardized diagnostic criteria it is impossible to establish a valid estimate of the pre-test probability of carpal tunnel syndrome

Method

In developing a diagnostic scale for carpal tunnel syndrome, a key consideration was that the potential users of this instrument included clinicians from a broad spectrum of backgrounds. It was hypothesized that focusing primarily on content experts, especially during the phase of item reduction, would reduce variation. It was thought that expert clinicians from different clinical fields would be more likely to agree than non-experts on the important diagnostic criteria. In addition to providing a greater likelihood of establishing a consensus, the use of opinion leaders was also considered critical to the eventual implementation of the scale on a broad basis. Participation of these individuals in the development of the instrument might facilitate its introduction to non-expert clinicians in the various disciplines.

Finally, commitment to the use of identified experts in the field of CTS would require the involvement of individuals from diverse geographic locations. An additional advantage to having an international perspective would be in reducing the risk of generating an instrument that was regional in scope and potentially dominated by concepts emanating from an influential local individual or institution.

To summarize, the main methodological considerations in accomplishing both item generation and reduction were: (i) consult diagnosticians from all the clinical disciplines involved in the care of CTS; (ii) focus on the recruitment of identified content experts on the topic of CTS, especially for the process of item reduction; (iii) minimize the risk of developing an instrument biased by concepts peculiar to a particular region by maintaining as broad a geographical perspective as possible.

Item generation

Items were generated from three main sources: literature review; in-depth interviews with local key informants; focus groups. A search of the MEDLINE database between 1966 and 1999 and of the HealthSTAR database from 1975 to 1999 was conducted using a variety of search terms including "CTS", "symptoms", "physical examination", "diagnosis" and "risk factors". Approximately 1500 articles were identified as having the diagnosis of CTS as their main focus. The abstracts of the retrieved titles were screened for relevance to the objective of item generation and the selected papers were fully reviewed.

In-depth key informant interviews were carried out with local content experts in the diagnosis of CTS. The content experts were, in most instances, identified from within the University of Toronto faculty. Individuals were selected for interview because they were recognized within their hospital community as experienced clinicians or because they held a specific perspective which was thought to be important to the process of assembling as complete a pool of potential items for inclusion in the final instrument as possible.

Focus groups were conducted at the national meetings of various specialties involved in the care of carpal tunnel syndrome. This setting allowed recruitment from a pool of participants from geographically diverse locations and different academic backgrounds. The model for organizing the groups was to obtain a list of pre-registrants to the event and to randomly select potential participants from among physicians attending the meeting. Prospective respondents were sent an information package describing the general aims of the study and the techniques of focus group interviewing. Telephone contact was then made with potential participants to solicit participation in the group. The group meeting to take place over a 90 minute period, including a luncheon, at the site of the meeting.

As a result of the literature review, the key informant interviews and the focus groups, a total of 57 unique items were identified (Appendix 3). These could be classified under six headings: patient characteristics (3); presenting symptoms (23); relationship between symptoms and work history (3); treatment associated with an improvement in symptoms (6); presence of coexisting medical conditions (8); physical examination findings (14). Many of these items were reported in all of the focus groups and key informant interviews.

A sample of Ontario clinicians listed in the specialties of neurology, neurosurgery, plastic surgery, orthopedic surgery, rheumatology, physiatry and occupational medicine were randomly selected from the Canadian Medical Directory using a random numbers table. Three hundred individuals were sent a package containing a list of the 57 items from the clinical history and physical examination. The participants were asked to answer the question: "How important is [item] in making a clinical diagnosis of CTS?" The responses were marked on a 10 cm visual analog scale with the anchors "completely unimportant" and "extremely important". The focus was on the clinical assessment and specifically excluded any mention of electrodiagnostic tests. Non-responders to our initial mailing were contacted a second time six weeks later. The study was closed six weeks after the second mailing.

The raw data was used to rank the importance assigned to each of the items by the participants. Cronbach's α was calculated for each specialty to measure the reliability of the responses. It was assumed that a high value of Cronbach's α would be indicative of the reliability of the sample. In other words, another sample selected in the same way, from the same population would be expected to give a similar result. This line of reasoning was used to evaluate the adequacy of sampling.

Cronbach's α for the raw data ranged between 0.80 and 0.92 for six of the seven clinical specialties. Cronbach's α for the whole sample was 0.97. This observation suggested that the reliability of the result was high and that sampling was adequate.

The intraclass correlation coefficient (ICC) was calculated for each specialty. These values ranged from 0.27 to 0.37 across the seven specialties. The ICC for the whole sample was 0.28.

	Ortho	Neurol	Rheum	Plast	Nsurg	Physiat	Occ	All
ICC	0.27	0.32	0.27	0.37	0.31	0.30	0.28	0.28
α	0.90	0.85	0.82	0.92	0.80	0.80	0.66	0.97

These findings showed that, as expected, there is poor consistency among Ontario specialists in assigning importance to items in the clinical evaluation. The consistency is equally low within and between specialties. This suggests that there are no clear specialty-specific diagnostic criteria.

Item reduction

Consistent with the overall goal of maintaining a geographically and clinically diverse basis for the development of the diagnostic criteria, a judgmental approach to item reduction, using Delphi as a group process, was selected. The absence of an obligation to meet in person greatly improves the feasibility of Delphi and lowers the cost significantly. There are no constraints on either the size or composition of the group and so the method lent itself to the development of a consensus among experts from a wide range of clinical backgrounds and working in different locations. The reliability of the group improves as the number of panelists is increased and the anonymous nature of the exercise insures that a single influential participant will not have a disproportionate effect on the outcome of the group. These factors were identified as key considerations in achieving a meaningful and useful consensus on the diagnostic criteria for CTS.

Criteria that indicate a consensus has been achieved will vary with the setting in which agreement is sought and the method being utilized. In this study, consensus was equated to a homogeneity of opinion expressed by a group of individuals rating items on a visual analog scale. In terms of measurement theory, homogeneity can be thought of as the capacity of an instrument to be consistent or reproducible. Cronbach's α is one statistical method, among others, which may be used to quantify the reliability of a summation of items. Where the items are highly correlated, they might be considered to be internally consistent or homogeneous. The interpretation of Cronbach's α is that the panelists chosen for the study are a sample from a population of content experts on the topic of CTS. An evaluation of the items by a new panel, chosen from this population, would have an expected correlation with the index sample equal to Cronbach's α . This line of reasoning justifies the use of Cronbach's α as the summarizing measure of the internal consistency for the assessment of all 57 items by the panel.

Expert panelists were recruited from as many of the pertinent clinical areas as possible including neurology [3], neurosurgery [2], rheumatology [2], occupational health [2], plastic surgery [2] and orthopaedic surgery [3]. Status as an expert was defined in the number of ways. In some cases the individual had an international reputation in the clinical field. This was especially true with respect to at least four of the surgeons and

one internist on the final panel. Other panelists were members of the American Society for Surgery of the Hand and the American Society for Peripheral Nerve. The remaining panelists had authored one or more papers directly concerned with CTS in a peer reviewed medical journal during the preceding three years.

A 10 cm visual analog scale was selected for evaluating each item. The scale was marked only with the anchors "completely unimportant" at the extreme left and "extremely important" to the right. The mean scores and standard deviations for each item were reported back to the respondents on the second iteration of the process. Following this second round, a consensus of opinion regarding the rankings appeared to have been achieved and the process was terminated.

Cronbach's α for the first round of the Delphi process was 0.86. The individual panelists - total group correlation ranged between 0.23 and 0.73 suggesting that there were some panelists who were relative outliers. Two of the three panelists with the lowest correlation with the entire group were from the same clinical specialty. The three highest values for correlation with the whole group were from a common clinical group as well.

Cronbach's α increased to 0.91 in the second iteration. The panelist-total group correlation was also substantially increased in every case as well. These values ranged from 0.58 to 0.75 corresponding to the higher Cronbach's α observed for the entire group. The disparity in correlation with the group among members of particular specialties was also markedly attenuated.

The qualitative meaning of this result is subject to interpretation and depends on the context however, reliability greater than 0.80 is normally considered substantial. Cronbach's α above 0.90 might be considered the minimum for a diagnostic scale.

Given that a consensus on the issue appeared to have been achieved, examination of the items in rank order suggested that the process was valid. The items favored by the panelists reflect the practices of many of the individuals interviewed in the item generation phase of the project in emphasizing basic history and physical examination findings in classic cases of CTS. Items with closely related meanings were also ranked together.

The most highly ranked items did not vary much between the two iterations. Each of the 20 highest ranked items could be grouped under one of five major domains: (i) distribution of sensory symptoms on history; (ii) physical evidence of motor denervation in the thenar eminence; (iii) nature of the sensory disturbance and features related to precipitating factors and exposures; (iv) physical signs of median nerve dysfunction; (v) the presence of coexisting medical conditions.

Rank	Iteration 1	Iteration2
1	Thenar atrophy	Sensory symptoms restricted to median nerve
2	Absence of sensory symptoms	Thenar atrophy
3	Weakness of APB	Sensory symptoms only outside median nerve
4	Sensory symptoms restricted to median nerve	Weakness of APB
5	Nocturnal sensory symptoms	Absence of sensory symptoms
6	Coexisting condition: Pregnancy	Nocturnal sensory symptoms
7	Coexisting condition: Diabetes mellitus	Sensory symptoms described as numbness or tingling
8	Sensory symptoms only outside median nerve	Coexisting condition: Pregnancy
9	Sensory symptoms described as numbness or tingling	Phalen sign
10	Phalen sign	Loss of two point discrimination
11	Precipitation by: Driving	Tinel sign
12	Loss of two point discrimination	Coexisting condition: Previous distal radius fracture
13	Coexisting condition: Hypothyroidism	Precipitation by: Driving
14	Decreased vibratory sensation	Decreased vibratory sensation
15	Response to: Splinting	Response to: Splinting
16	Tinel sign	Response to: Steroid injection
17	Coexisting condition: Previous distal radius fracture	Precipitation by: grasping the newspaper
18	Exposure to vibration	Exposure to vibration
19	Coexisting condition: Inflammatory arthropathy	Coexisting condition: Hypothyroidism
20	Dropping objects	Precipitation by: grasping the phone

The rankings of the items provided some interesting insights into the importance that the experts attached to a number of clinical findings frequently discussed in the literature on CTS. The most highly ranked items were ones most usually included in descriptions of the classic presentation of CTS. Sensory symptoms described as numbness or tingling, restricted to the median nerve distribution and which frequently interrupt sleep, were highly rated by the panel as diagnostic features in the history. Physical examination findings like weakness and atrophy of the abductor pollicis longus muscle, a loss of two-point discrimination, the Phalen test and Tinel sign were also emphasized by the experts.

The panel placed smaller emphasis on a number of items frequently associated with CTS. Work-relatedness received relatively little attention as a diagnostic criterion. For example, worsening of symptoms by exposure to computer keyboard work was ranked in the lower half of items at 31st. Worsening of symptoms by repetitive work [unspecified],

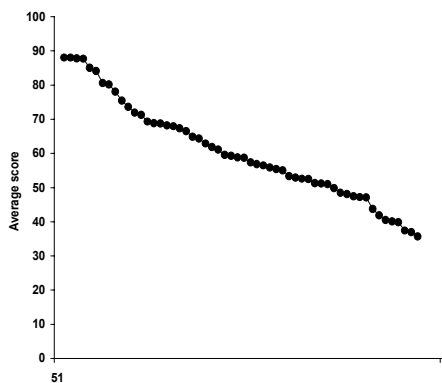
other than computer keyboarding, was ranked slightly higher at 21st. A change in symptoms with modifications in work activity were ranked 27th as a diagnostic item. Obesity, frequently described as coexisting with CTS, was ranked 49th indicating that this finding plays little or no role in the evaluation of experts. Similarly, gender [ranked 37th] and patient age [50th] have little bearing on the diagnosis made by the panelists.

These results point out some interesting issues with respect to the way in which the literature may influence the establishment of diagnostic criteria. Even the demonstration of a strong association or correlation between a diagnosis like CTS and another factor like obesity obviously may be entirely spurious and does not necessarily indicate causality or any other kind of relationship. However, the implicit interpretation of data of this nature is frequently that there is some kind of non-random connection between the two observations with the implication that the presence of one can be predicted, to a certain extent, by the observation of the other. The non-expert may attach unwarranted diagnostic significance to this kind of relationship.

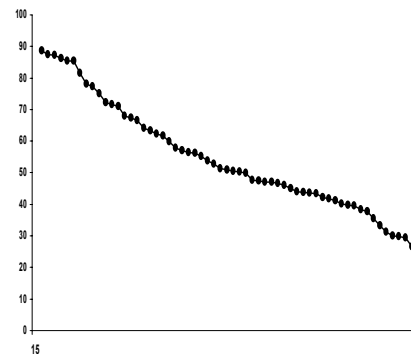
Inclusion of items in the final scale

A number of methods were considered for the final stage of item reduction. The average scores of the items, ranked in descending order was plotted in order to determine if there was a sudden change in slope, or inflection point, at one or more locations. This might suggest a natural truncation point reflective of a clear distinction made by the panelists between the items on either side of this division. After both iterations the largest slope among the twenty highest rated factors was between items was between the sixth and seventh items.

Iteration 1



Iteration 2



Although a formal factor analysis was not performed, examination of the twenty highest ranked items after the second Delphi iteration showed that several of these could be combined together. For example, the items ranked 1,3,5 and 7 all related to the nature and distribution of the sensory disturbance. Items ranked 2 and 4 described denervation of the thenar musculature. Items ranked 8, 12 and 19 were related to co-existing

conditions medical conditions. Items 15 and 16 described the response to common therapeutic interventions for CTS. On this basis, the list of items for inclusion was decreased to eight:

- Numbness in the median nerve distribution
- Nocturnal sensory symptoms
- Ameliorating /exacerbating factors
- Co-existing medical conditions: pregnancy, hyperthyroidism, diabetes

- Tinel sign
- Phalen test
- Loss of two-point discrimination
- Weakness of abductor pollicis brevis

Weighting and validation of items in the scale

The objective was to develop a logistic regression model with the dependent variable reflecting the probability of CTS and the independent variables, the items in the diagnostic scale. Case histories can be constructed to contain specific levels of the variables under consideration. For this diagnostic index, all of the items were binary rather than continuous or ordinal variables. Consequently, profiles were constructed to reflect the presence or absence of the various items rather than a specific level.

The number of possible combinations for eight binary factors is 2^8 or 256. A unique case history was created for each of these combinations. The case histories had a standardized format that indicated a patient age, gender and occupation. These were randomly assigned to the case history except for cases where age or gender was pertinent, for example in the case of pregnancy as a diagnostic factor.

The items “ameliorating/exacerbating factors” and “co-existing conditions” represented a consolidation of a few factors in each case. These were randomly distributed among the case histories in which items occurred. For example, “co-existing conditions” included pregnancy, hypothyroidism and diabetes mellitus. These were each inserted in one-third of the profiles in which the item “co-existing conditions” was being tested.

Finally, an equal number of case histories representing non-CTS patients were also created. The conditions that were selected are commonly seen in the same patient population presenting with CTS or are frequently misdiagnosed as CTS. These included: cubital tunnel syndrome, de Quervain’s tenosynovitis, trapeziometacarpal joint osteoarthritis, lateral epicondylitis, flexor carpi radialis tendonitis, flexor tendon triggering, radial sensory nerve compression and osteoarthritis or the wrist.

A total of 512 case histories were created representing 256 cases of CTS and 256 non-CTS cases. Two new expert panels were recruited to evaluate the case histories. One panel rated the probability, on a 10 cm visual analog scale, that the case history represented CTS. The second panel evaluated the same case histories and were asked to

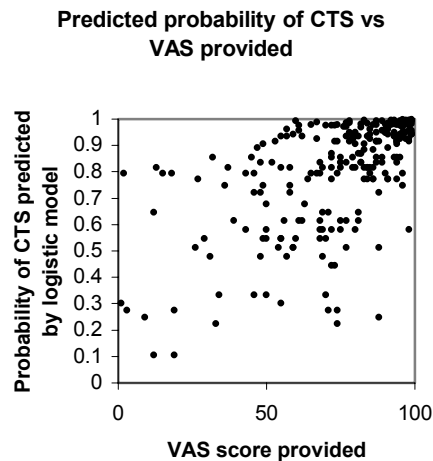
make a binary judgment as to whether the case represented carpal tunnel syndrome. Each panel comprised 16 individuals who each evaluated 32 case histories.

The cases sent to each panelist included equal numbers of randomly selected CTS and non-CTS case histories. In order to minimize the risk that rater differences in the use of the visual analog scale would affect the value placed on any of the items, the cases were allocated to each rater so that all features were uncorrelated and equally represented within each set of cases. This approach also ensured that within one rater, one factor would not spuriously increase the importance of another factor through inadvertent correlations.

The panelists participating in this phase of the study did not include individuals who participated in the focus groups or in the preceding Delphi process. The panelists were clinicians who regularly cared for individuals with CTS. They were similar to those participating in the Delphi process with respect to the diversity of clinical backgrounds and geographic location of practice however they did not necessarily meet the criteria for being identified as experts, defined for the Delphi panelists. The reasons for this were that the pool of clinicians meeting these criteria was too small to make the use of only these individuals for the validation phase feasible. Secondly, the use of less rigid criteria for being included as a panelist would be more likely to maximize the generalizability of the final instrument.

Results

Only six of the eight factors considered for inclusion contributed to the explanatory power of the model. The last two features, “ameliorating and exacerbating factors” and “co-existing medical conditions” were not included in the final instrument. The correlation between the probability of carpal tunnel syndrome predicted by the model and the probability assigned to each case by an expert was 0.60.



Conversion of the raw visual analog scale (VAS) data to z-scores increased the correlation to 0.68 by reducing the variability in the VAS responses. The values generated by the logistic regression model were highly correlated ($r = 0.91$) with values for binary diagnosis (where $p > 0.5$ indicates CTS present) predicted linear regression

models. In addition, the correlation between the values predicted by the multiple regression and logistic regression models were also very high, around 0.99. These findings suggest all three models were highly consistent with one another in the description of the data although only the logistic regression model was tested against an independent data set. The logistic regression model also provided the best range of potential values for the probability of CTS.

Conclusions

1. There is substantial variability in the way in which the clinical diagnosis of CTS is established by clinicians working in disciplines commonly involved in the care of this condition.
2. Consensus among experts can be achieved through the use of anonymous group processes like Delphi. In this study, Delphi was particularly useful because it allowed the inclusion of panelists working in a range of clinical settings in geographically divergent locations. Cronbach's α was a useful statistic for expressing the extent of consensus within panel. A convincing degree of agreement as to the relative importance of clinical items for the diagnosis of CTS was observed.
3. The most important items used for the clinical diagnosis of CTS can be weighted to increase the correlation between the probability of CTS predicted by the model and that assigned by a clinical expert. A logistic regression model is best suited to this objective because it expresses the dependent variable as a probability and allows a feasible range of potential values subject to a minimal floor effect. The model may be more accurate for identifying the diagnosis of CTS than for ruling it out.