Development and Validation of Diagnostic Criteria for Carpal Tunnel Syndrome

Brent Graham, MD, Glenn Regehr, PhD, Gary Naglie, MD, James G. Wright, MD

From the University Health Network, Hospital for Sick Children, Toronto Rehabilitation Institute, Toronto, Ontario, Canada; and the University of Toronto, Toronto, Ontario, Canada.

Purpose: To develop clinical diagnostic criteria for carpal tunnel syndrome (CTS) that modeled the clinical diagnostic practices of experts.

Methods: Fifty-seven clinical findings associated with CTS had been ranked previously in order of diagnostic importance using Delphi as a method of establishing consensus among a panel of expert clinicians. The 8 most highly ranked criteria then were placed into all possible combinations to create 256 unique case histories. Two new panels of experts rated these case histories. One panel made a binary evaluation as to whether the case history did or did not represent CTS. This allowed the development of a logistic regression model that had the probability of carpal tunnel syndrome as the dependent variable and the weighted diagnostic criteria as the independent variables. This model then was validated against the judgments of the second panel of clinicians who estimated the probability of CTS for each of the same case histories.

Results: The correlation between the probability of CTS predicted by the model and the panel of clinicians was 0.71.

Conclusions: The most important clinical diagnostic criteria for CTS as identified from a larger pool of potential diagnostic items through a consensus approach using Delphi were weighted and found to correlate well with the judgments of a new panel of clinicians. By improving the consistency of the diagnosis of CTS these criteria should lead to more effective treatment and a better understanding of the effect of workplace exposures in the development of this condition. A methodology that emphasizes a rigorous approach to item generation and item reduction through expert consensus, followed by validation, may represent a template for establishing consensus among experts on other controversial clinical issues. (J Hand Surg 2006;31A:919.e1–919.e7. Copyright © 2006 by the American Society for Surgery of the Hand.)

Type of study/level of evidence: Diagnostic, Level I.

Key words: Carpal tunnel syndrome, consensus, diagnosis, logistic regression.

eonly **Additional material is available online.**

Carpal tunnel syndrome (CTS) is diagnosed commonly, with prevalence estimates in the general population varying from 0.1% to 9.2%.[1–5] This variation in prevalence estimates is caused in part by different diagnostic criteria for CTS.[1,6–10] Misdiagnosis, which probably is frequent, may lead to the inappropriate use of electrodiagnostic tests and, more importantly, inappropriate treatment. The inaccurate diagnosis of CTS has been identified as one of the most common causes of treatment failure for CTS.[11,12]

There is no gold standard for establishing a diagnosis of CTS. The diagnosis of CTS has been based on both clinical findings[8–10] and the results of electrodiagnostic testing.[6,7,13] Although electrodiagnostic studies often are assumed to represent a diagnostic gold standard for CTS,[14–16] this assumption may be questioned for at least 3 reasons. First, because electrodiagnostic study results are abnormal only when compression is sufficiently severe to cause structural abnormalities in the median nerve,[17,18] nerve conduction test results may be normal despite the presence of clinically significant median nerve compression. Thus improvement is observed in many patients after surgery for CTS despite normal electrodiagnostic test results.[19,20] Second, it is assumed that nerve conduction velocities are distributed normally with thresholds for abnormality defined as the lower 2.5% to 5% of values; however, the distribution of nerve conduction velocities is skewed substantially toward slower conduction velocities in the asymptomatic population,[21,22] resulting in a large proportion of healthy individuals being mislabeled as affected by CTS. Third, the cut-off point used for determining abnormality in the measurement of nerve conduction is highly variable.[14,23,24]

Some clinical experts rely exclusively on clinical findings and do not use electrodiagnostic study results to make the diagnosis of CTS.[25] Thus symptoms and signs play a major role in establishing the diagnosis of CTS. Clinicians, however, are inconsistent in the importance they ascribe to the various clinical findings.[26] Therefore an important step in reducing inconsistencies and misdiagnoses would be the development of standardized clinical criteria for CTS that could be applied consistently. The objective of this study was to establish consensus among clinical experts on the most important clinical diagnostic criteria for CTS, give these criteria weighted values, and then validate the resulting diagnostic instrument.

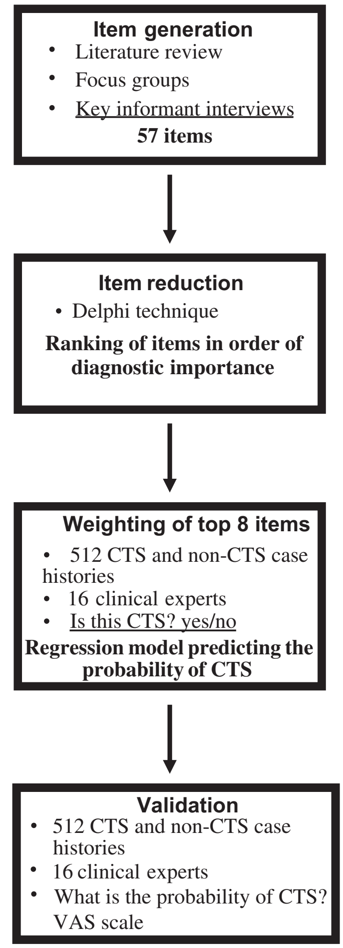

Figure 1. Summary of the steps in the development of a diagnostic scale for CTS.

Materials and Methods

A diagnostic scale for CTS was developed in 3 stages (Fig. 1). For the process of item generation 57 potential diagnostic criteria for CTS were identified through a combination of literature review, key informant interviews, and focus groups with clinicians from various clinical backgrounds.[27] This pool of potential diagnostic items was rated for importance by a panel of clinicians who met our criteria for being experts in the clinical diagnosis of CTS. An individual was considered an expert if he or she had published reports in peer-reviewed journals on the topic of CTS management or had a national reputation in the field of peripheral nerve conditions. Ranking of these criteria in order of diagnostic importance thus was established using the Delphi technique, a process that allowed a group consensus to be established through repeated assessments of the criteria.[28]

The decision about the number of items from this list that should be included in a final diagnostic scale required a balance between comprehensiveness and feasibility. Examination for commonality and redundancy among the 20 highest-ranked items showed that several of these could be combined together. For example, the items ranked 1, 3, 5, and 7 in importance 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 coexisting medical conditions. Items 15 and 16 described the response to common therapeutic interventions for CTS. This process of combining similar clinical items to create broader diagnostic constructs allowed the top 20 items to be collapsed into 8 major criteria (Table 1).

The relative importance of the final 8 items was determined by establishing weights for each criterion. The objective was to develop a logistic regression model with the dependent variable being the probability of CTS and the independent variables being the items in the diagnostic scale. This process required 2 new groups of 16 clinicians each. None of these individuals had participated in any of the earlier stages of the project.

These clinicians examined case histories containing the 8 diagnostic criteria in all possible combinations.[29,30] This required the creation of 256 unique case histories. This approach was taken so that all of the 8 criteria could be evaluated as they occurred alone in a case or together with the other criteria. Another 256 non-CTS case histories also were created to present the clinicians with conditions that are seen commonly in the same patient population presenting with CTS or that frequently are misdiagnosed as CTS. These included cases representing cubital tunnel syndrome, de Quervain's tenosynovitis, trapeziometacarpal joint osteoarthritis, flexor tendon triggering, osteoarthritis of the wrist, lateral epicondylitis, flexor carpi radialis tendinitis, and radial sensory nerve compression. Each case history had a standardized format that included a randomly assigned patient age, gender, and occupation except in instances in which age or gender was intrinsic to the item (eg, pregnancy as a diagnostic factor). Two typical case histories are shown in Appendix A (Appendix A may be viewed at the Journal's Web site, www.jhandsurg.org).

Each of the 32 clinicians evaluated 16 randomly selected CTS and 16 non-CTS case histories. 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 1 rater, 1 factor would not increase spuriously the importance of another factor through inadvertent correlations. This complicated approach was necessary to the goal of developing an unbiased logistic regression model to predict the probability of CTS. The use of actual patients rather than case histories would have introduced a number of uncontrollable variables that would have thwarted this objective. The use of vignettes such as those used in our study has been shown to be more effective than histories obtained from patient charts.[31]

One panel of 16 clinicians made a binary judgment as to whether or not each case represented CTS. These data were used to develop a logistic regression model with the dependent variable being the probability of CTS and the predictor variables being the 8 diagnostic criteria. A second panel of 16 clinicians who evaluated the same case histories rated the probability of CTS for each case history on a 10-cm visual analog scale (VAS). These ratings were compared with the predictions of the logistic regression model. The process of reviewing the case histories is summarized in Figure 2.

Table 1. Final List of Unweighted Clinical Diagnostic Criteria

Numbness and tingling in the median nerve distribution
Nocturnal numbness
Weakness and/or atrophy of the thenar musculature
Tinel's sign
Phalen's test
Loss of 2-point discrimination
Ameliorating/exacerbating factors*
Improvement by splinting and/or steroid injection
Worsening with activities such as driving and strenuous hand use
Coexisting medical conditions*
Pregnancy
Diabetes
Hypothyroidism

*These factors did not contribute significantly to the logistic regression model.

Results

Only 6 of the 8 criteria (numbness in the median nerve distribution, nocturnal numbness, weakness/atrophy of the thenar musculature, Tinel's sign, Phalen's test, loss of 2-point discrimination) contributed significantly ($p < .05$) to the model (Appendix B may be viewed at the Journal's Web site, www.jhandsurg.org). By applying the resulting formula to each of the cases a CTS probability was predicted for each case by the model.

The probability of CTS for each case history predicted by the model then was compared with the probability of CTS independently assigned to each case by the clinicians using the VAS. The correlation between the probability of CTS predicted by the model and that estimated by the clinical experts was 0.71 (Fig. 3).

Discussion

A failing of contemporary clinical medicine in general is the lack of clearly stated, widely agreed on, and formally established diagnostic criteria for many, if not most, medical conditions. Diagnosis, an essential component of clinical decision making, often determines which patients will be treated and what treatment they will receive. Thus standardization of diagnostic practices for commonly diagnosed conditions is of fundamental importance to the practice of clinical medicine.

The term *gold standard*, which often is used in clinical medicine, implies a criterion for definitive diagnosis; however, gold standards exist only by consensus. If there is no consensus on the diagnostic criteria for a condition then a standard of diagnosis

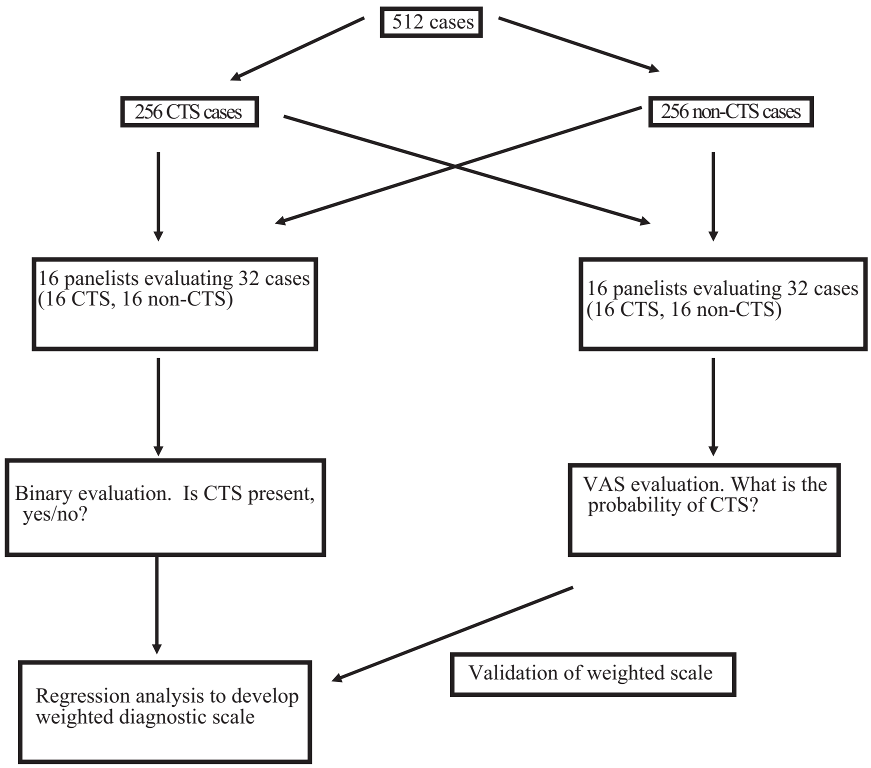

Figure 2. Schema for dstribution of cases for weighting criteria in the final scale.

cannot be established. The existence of agreement is the key issue. This is true even when a tissue diagnosis is made because unless there is agreement on the meaning of the histologic findings the significance remains in doubt. In the diagnosis of CTS electrodiagnostic tests often are used as the gold standard criterion for this condition. There is no consensus on this issue, however, and thus electrodiagnostic tests cannot be considered a gold standard criterion for the diagnosis of CTS. This is not to assert that electrodiagnostic tests never play a role in the diagnosis of CTS but rather to suggest that they cannot on their own be considered the criterion for the diagnosis of CTS in all circumstances, which is necessary if these investigations are to be considered the gold standard.

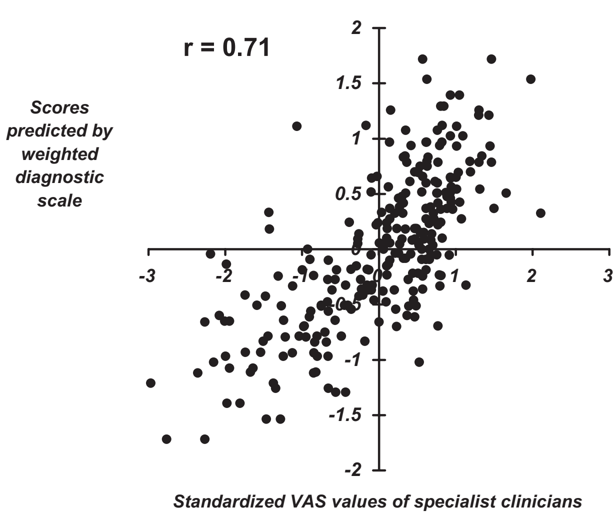

Figure 3. Standardized VAS scores from specialist clinicians versus scores predicted by the weighted diagnostic scale. r = 0.71.

Our study has developed standardized clinical diagnostic criteria for CTS that have been validated against the opinions of clinical experts. The correlation with the clinical experts of 0.71 was substantial, explaining more than 50% of the variance in the sample. Standardized diagnostic criteria such as those established by our study have many uses. First, standard criteria could be used to determine the population prevalence of CTS. The use of valid standardized diagnostic criteria for CTS may decrease the wide variance in population prevalence caused by variable case definitions. This has been attempted in the past[32] but the case definitions intended for large-scale epidemiologic studies often are not applicable for clinical use in individuals. The study of etiologic factors such as workplace exposures would be aided by the adoption of standardized diagnostic criteria because variability in the identification of CTS would be limited.

Second, nonexpert clinicians such as those in primary care settings could use standardized criteria to establish the probability of CTS in a manner approaching the practices of clinical experts. Depending on the probability of CTS the action taken might be the initiation of nonsurgical treatment, referral for surgery, further investigation with electrodiagnostic testing, or consideration of an alternative diagnosis. For example, if the probability of CTS is high then treatment with either nonsurgical or surgical interventions probably can be instituted safely without electrodiagnostic testing. At a high level of clinical probability for CTS electrodiagnostic tests are not likely to add much greater certainty to the diagnosis. When the probability of CTS is very low it may be appropriate to consider a different diagnosis. Even when electrodiagnostic test results are positive in this group the posttest probability of CTS still is likely to be insufficient to recommend surgical intervention. Estimates of probability that are intermediate may indicate a need for further investigation with electrodiagnostic testing. This is the group in whom the results of electrodiagnostic testing are most likely to influence decision making about treatment. Furthermore electrodiagnostic test data could be interpreted in a Bayesian context using the output of the diagnostic instrument as an estimate of pretest probability. The ability to report the electrodiagnostic test results as a posttest probability may improve their precision beyond the current standard of establishing the CTS diagnosis based on comparison with a threshold for nerve conduction velocity.

Third, standardized diagnostic criteria also could be used to establish guidelines for instituting treatment such as surgery. For example surgical decompression of the carpal tunnel might be considered if the probability of CTS exceeds a defined threshold value. This threshold might vary depending on the clinical context—for example if there has been treatment that failed or if the patient is receiving workers' compensation benefits. The decision to institute treatment is based on the severity of the condition to be treated and a variety of factors associated with treatment including the risk for complications, morbidity, and cost. As such it is difficult to determine precisely a critical threshold at which surgical treatment should be recommended. The criterion used to make important clinical decisions should have a reliability of at least 0.80.[33] It might be considered that a CTS probability of at least this level might be necessary to start treatment. A determination of the actual threshold for recommending different forms of treatment requires further study; however, the first step is to base the diagnosis of CTS on standardized criteria so that these comparisons can be made.

Our study has some limitations. Because they are based on expert opinion these diagnostic criteria may be relevant only to current thinking on the topic of the clinical diagnosis of CTS. As knowledge grows the views of experts may change through time and as a result an instrument based on expert opinion may become obsolete without constant re-evaluation. The *Diagnostic and Statistical Manual of Mental Disorders IV*[34] is an example of an expert-based diagnostic system that has had successive revisions aimed at capturing the evolution of the concepts of mental disease. Although the diagnostic criteria for conditions in the *Diagnostic and Statistical Manual of Mental Disorders* have changed over time, they have been based consistently on a consensus of experts. In general although the nature of a consensus may evolve, the principle of obtaining consensus should remain the key consideration in establishing diagnostic criteria. It may be necessary to determine periodically whether expert consensus is changing. As evidence linking outcomes with these diagnostic criteria accumulates the reliance on expert consensus probably will decrease.

A second limitation of our diagnostic instrument is that the validation process was performed using simulated case histories created for this purpose. Although the use of clinical vignettes is a more valid approach than the abstraction of clinical data from patient charts,[31] the instrument requires testing in actual patients.[35] The reliability of some of the individual clinical criteria is unknown and may be variable in different settings; however, the output of the instrument, which expresses the diagnosis in probabilistic terms, allows for flexibility on the part of the clinician. In other words the clinician may readjust the probability predicted by the instrument if aspects of the history or physical examination for CTS are equivocal or if there are other factors not included in the instrument that are thought to have diagnostic implications. These questions can be answered only in studies on patients. These studies currently are underway in our center.

We have developed clinical diagnostic criteria for CTS based on expert consensus. The diagnostic instrument correlates well with the judgments of expert clinicians. The diagnostic instrument could be used to standardize the clinical diagnosis of CTS and this should increase the accuracy of diagnosis, improve the targeting of specific treatments, and facilitate the performance of epidemiologic studies of the risk factors for CTS. The role and interpretation of electrodiagnostic test results in the diagnosis of CTS may be defined better as a result of the use of standardized diagnostic criteria. We believe that a formal and methodologically sound approach to developing diagnostic criteria similar to that used in this study could be used in other contexts when consensus on diagnosis is lacking.

Received for publication November 4, 2005; accepted in revised form March 14, 2006.

Supported by the Physicians' Services Incorporated Foundation (grant 97-52), the Mary Trimmer Chair in Geriatric Medicine Research, University of Toronto (G.N.), and the R.B. Salter Chair of Surgical Research and an Investigator of the Canadian Institute for Health Research (J.G.W.).

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

Corresponding author: Brent Graham, MD, Toronto Western Hospital, 399 Bathurst St, FP-174, Toronto, Ontario M5T 2S8, Canada; e-mail: brent.graham@utoronto.ca.

0363-5023/06/31A06-0008$32.00/0
doi:10.1016/j.jhsa.2006.03.005

References

1. Katz JN, Larson MG, Sabra A, Krarup C, Stirrat CR, Sethi J, et al. The carpal tunnel syndrome: diagnostic utility of the history and physical examination findings. Ann Intern Med 1990;112:321–327.
2. Levine DW, Simmons BP, Koris MJ, Daltroy LH, Hohl GG, Fossel AH, Katz JN. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. J Bone Joint Surg 1993;75A:1585–1592.
3. Atroshi I, Gummesson C, Johnsson R, Ornstein E, Ranstam J, Rosen I. Prevalence of carpal tunnel syndrome in a general population. JAMA 1999;282:153–158.
4. de Krom MCTFM, Kester ADM, Knipschild PG, Spaans F. Risk factors for carpal tunnel syndrome. Am J Epidemiol 1990;132:1102–1110.
5. Tanaka S, Wild DK, Seligman PJ, Behrens V, Cameron L, Putz-Anderson V. The US prevalence of self-reported carpal tunnel syndrome: 1988 National Health Interview Survey data. Am J Public Health 1994;84:1846–1848.
6. Katz JN, Larson MG, Fossel AH, Liang MH. Validation of a surveillance case definition of carpal tunnel syndrome. Am J Public Health 1991;81:189–193.
7. Katz JN, Simmons BP. Clinical practice. Carpal tunnel syndrome. N Engl J Med 2002;346:1807–1812.
8. Cummings K, Maizlish M, Rudolph L, Dervin K, Ervin A. Occupational disease surveillance: carpal tunnel syndrome. MMWR Morb Mortal Wkly Rep 1989;38:485–489.
9. Margolis W, Kraus JF. The prevalence of carpal tunnel syndrome symptoms in female supermarket checkers. J Occup Med 1987;29:953–956.
10. Silverstein BA, Fine LJ, Armstrong TJ. Occupational factors and carpal tunnel syndrome. Am J Ind Med 1987;11:343–358.
11. Hunt TR, Osterman AL. Complications of the treatment of carpal tunnel syndrome. Hand Clin 1994;10:63–71.
12. Kessler FB. Complications of the management of carpal tunnel syndrome. Hand Clin 1986;2:401–406.
13. de Krom MCTFM, Knipschild PG, Kester ADM, Spaans F. Efficacy of provocative tests for diagnosis of carpal tunnel syndrome. Lancet 1990;335:393–395.
14. Nathan PA, Keniston RC, Meadows KD, Lockwood RS. Predictive value of nerve conduction measurements at the carpal tunnel. Muscle Nerve 1993;16:1377–1382.
15. Boniface SJ, Morris I, MacLeod A. How does neurophysiologic assessment influence the management and outcome of patients with carpal tunnel syndrome? Br J Rheum 1994;33:1169–1170.
16. Golding DN, Rose DM, Selvarajah K. Clinical tests for carpal tunnel syndrome: an evaluation. Br J Rheum 1986;25:388–390.
17. Campion D. Electrodiagnostic testing in hand surgery. J Hand Surg 1996;21A:947–956.
18. Wilcox MS, Bilbao A. Sensitivity of electrophysiological studies and the carpal tunnel syndrome. Muscle Nerve 1993;16:1265–1266.
19. Braun RM, Jackson WJ. Electrical studies as a prognostic factor in the surgical treatment of carpal tunnel syndrome. J Hand Surg 1994;19A:893–900.
20. Grundberg A. Carpal tunnel decompression in spite of normal electromyography. J Hand Surg 1983;8:348–349.
21. Goodgold J. A statistical problem in diagnosis of carpal tunnel disease. Muscle Nerve 1994;17:1490–1491.
22. Robinson LR, Temkin NR, Fujimoto WY, Stolov WC. Effect of statistical methodology on normal limits in nerve conduction studies. Muscle Nerve 1991;14:1084–1090.
23. Jackson DA, Clifford JC. Electrodiagnosis of mild carpal tunnel syndrome. Arch Phys Med Rehabil 1989;70:199–204.
24. Stevens JC. AAEE minimonograph #26: the electrodiagnosis of carpal tunnel syndrome. Muscle Nerve 1987;10:99–113.
25. Duncan KH, Lewis RC, Foreman KA, Nordyke ME. Treatment of carpal tunnel syndrome by members of the American Society for Surgery of the Hand: results of a questionnaire. J Hand Surg 1987;12A:384–391.
26. Graham B, Dvali L, Regehr G, Wright JG. Variations in diagnostic criteria for carpal tunnel syndrome among Ontario specialists. Am J Ind Med 2006;49:8–13.
27. Graham B. The development of diagnostic criteria for carpal tunnel syndrome. MSc: University of Toronto, 2003:52–60.
28. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. J Clin Epidemiol 2003;56:1150–1156.
29. Gustafson DH, Fryback DG, Rose JH, Yick V, Prokop CT, Detmer DE, Moore J. A decision theoretic methodology for severity index development. Med Decis Making 1986;6:27–35.
30. Fishburn WR. Experiences-in-being. J Am Coll Health Assoc 1970;18:288–291.
31. Peabody JW, Luck L, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients and chart abstraction: a prospective validation of 3 methods for measuring quality. JAMA 2000;283:1715–1722.
32. Rempel D, Evanoff B, Amadio PC, de Krom M, Franklin G, Gray R, et al. Consensus criteria for the classification of carpal tunnel syndrome in epidemiologic studies. Am J Public Health 1998;88:1447–1451.
33. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. BMJ 1997;314:572–573.
34. American Psychiatric Association Task Force on DSM-IV. Diagnostic and statistical manual of mental disorders: DSM-IV. 4th ed. Washington, DC: American Psychiatric Association, 1994.
35. Jones TV, Gerrity MS, Earp J. Written case simulations: do they predict physician behaviour? J Clin Epidemiol 1990;43:805–815.

Appendix A. Two Typical Case Histories Provided to the Panelists

Patient number 522 is a 32-year-old male who presents with a chief complaint of numbness/tingling in the distribution of the median nerve. There are no other symptoms. The symptoms are made worse by grasping. There has been some improvement with splinting.

The past history includes no other relevant conditions.

The physical examination shows a positive Phalen's test and a Tinel's sign over the median nerve at the level of the carpal tunnel The remainder of a comprehensive physical examination is within normal limits.

On the basis of only the clinical information provided, does this patient has carpal tunnel syndrome?

The rater giving a binary response classified this as *yes*.

The rater responding on the VAS scored the history as *76*.

Case number 773 is a 63-year-old male who presents with a chief complaint of numbness and tingling in the ring and small fingers.

The past history includes hypothyroidism. The patient is retired.

The physical examination shows increased numbness with passive flexion of elbow, positive Tinel's sign over cubital tunnel, and normal intrinsic muscle strength. The remainder of a comprehensive physical examination is within normal limits.

On the basis of only the clinical information provided, estimate the probability that this patient has carpal tunnel sydrome?

The rater giving a binary response classified this as *no*.

The rater responding on the VAS scored this history as *12*.

Appendix B. Logistic Regression Formula for the 6-Item Model Predicting the Probability of CTS

$p(CTS)/1 - p(CTS) = e^{b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6}$ where $b_0 = -2.14$; variable present = 1, absent = 0; $b_1 = 1.44$; x_1 = thenar atrophy; $b_2 = 1.44$; x_2 = Phalen's test; $b_3 = 1.30$; x_3 = loss of 2-point discrimination; $b_4 = 1.16$; x_4 = Tinel's sign; $b_5 = 1.16$; x_5 = nocturnal numbness; $b_6 = 1.03$; x_6 = numbness, median nerve distribution.