

# A Bayesian argument against rigid cut-offs in electrodiagnosis of median neuropathy at the wrist

H. Nodera, MD; D.N. Herrmann, MB, BCh; R.G. Holloway, MD, MPH; and E.L. Logigian, MD

**Abstract—Background:** Nerve conduction (NC) tests, using rigid cut-offs separating normal from abnormal test values, are commonly used to confirm median neuropathy at the wrist (MNW). The authors studied patients with clinically defined mild MNW and a normal median distal motor latency to determine 1) how much sensory or mixed NC test results increase (or decrease) the probability of MNW and 2) the NC test values required to confirm (or exclude) MNW for the range of pretest probabilities of MNW. **Methods:** Palmar, digit 4 (D4), and digit 2 (D2) median NC tests were reviewed in 125 hands with mild carpal tunnel syndrome (CTS) and 100 control hands with musculoskeletal pain. Receiver operating characteristic curves and interval likelihood ratios were plotted for the three tests. Using Bayes theorem, post-test probability of MNW was then determined for the range of pretest probabilities and NC test values. **Results:** Receiver operating characteristic curves showed that for a set specificity of 97%, palmar and D4 studies had higher electrodiagnostic utility than D2 studies with cut-off test values (sensitivities of 0.3 msec, 64.0%; 0.4 msec, 71.2%; and 50 m/sec, 44.8%). However, Bayesian analysis showed that to confirm MNW more conservative cut-off values (palmar 0.5 msec, D4 0.7 msec, D2 44 m/sec) were required for pretest probabilities  $\leq 50\%$ , whereas borderline abnormal values (palmar 0.4 msec, D4 0.5 msec, D2 48 m/sec) sufficed when pretest probabilities were  $\geq 75\%$ . Conversely, normal test values could exclude MNW only for pretest probabilities  $< 25\%$ . **Conclusions:** For a given NC test value, post-test probability of MNW can be determined from the estimated pretest probability (derived from clinical data), interval likelihood ratios, and Bayes theorem. Use of rigid cut-off values to confirm MNW is problematic, because more conservative cut-offs are required for low pretest probability. Conversely, NC tests with sensitivity  $< 95\%$  cannot exclude MNW when pretest probability is high.

NEUROLOGY 2003;60:458–464

Nerve conduction (NC) tests are commonly used in patients with symptoms of carpal tunnel syndrome (CTS)<sup>1,2</sup> to confirm the presence of median neuropathy at the wrist (MNW).<sup>3,4</sup> NC evaluation of patients with possible MNW has evolved over the years to increase diagnostic sensitivity by shortening the median nerve segment tested (e.g., palmar median study)<sup>5-7</sup> or by using an internal control, ulnar nerve segment for comparison (e.g., digit 4 [D4] median vs D4 ulnar),<sup>4,8,9</sup> or both (e.g., palmar median vs ulnar comparison).<sup>4,10</sup> High electrodiagnostic sensitivity and specificity is particularly important in the subgroup of patients with mild CTS in whom results of routine studies are often normal, and in whom the palmar and D4 median–ulnar comparison NC tests are often required.

However, most reports of these NC techniques have set normal cut-off values, thereby dichotomizing the continuous test value range into normal and abnormal. There are several problems with this approach. First, there are no ideal cut-off values that yield 100% sensitivity and 100% specificity. Setting cut-off values requires a trade-off of false-positives for false-negatives: the lower the one, the higher the other. Second, classifying test values as either nor-

mal or abnormal loses information, because an extremely abnormal test value is more likely to be a true positive than is a mildly abnormal result. Third, this standard approach does not account for variation in pretest probability (PreTP) of disease as put forth in Bayes theorem; that is, that the post-test probability (PostTP) of a specific abnormal test result declines as PreTP of disease declines.<sup>11-13</sup> Thus, when PreTP of disease is low, an abnormal test result using the standard cut-offs of mean  $\pm 2$  SD, for example, may not suffice to confirm MNW.

These problems derive from viewing the diagnostic impact of an electrophysiologic test solely on its sensitivity and specificity for a rigid cut-off value. In this report, we focus on PostTP of disease as determined by the actual test value obtained and hypothetical estimates of PreTP of disease. Using patients with clinically defined mild CTS (and with a normal median distal motor latency) and a control group with musculoskeletal pain, the purpose of this study was to evaluate three commonly performed NC tests used in the electrodiagnostic evaluation of patients with putative CTS. For each of the three tests, our goals were to determine the following: 1) the PostTP of MNW for all NC test values and PreTP, 2)

From the Department of Neurology, University of Rochester School of Medicine and Dentistry, Rochester, NY.

Received May 24, 2002. Accepted in final form October 3, 2002.

Address correspondence and reprint requests to Dr. Eric Logigian, Box 673, Department of Neurology, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, NY 14642; e-mail: eric\_logigian@urmc.rochester.edu

458 Copyright © 2003 by AAN Enterprises, Inc.

**Table** Summary of demographic and electrophysiologic data

Characteristics	Control	CTS	<i>p</i> Value	Sensitivity, %	Specificity, %
Number of hands (patients)	100 (86)	125 (112)			
Male/female	22/64	21/91	NS*		
Age, y, mean $\pm$ SD (range)	41.2 $\pm$ 10.8 (14–68)	44.2 $\pm$ 12.3 (21–85)	NS†		
<b>Palmar</b>					
Onset difference, msec (normal $\leq$ 0.3); no. of abnormal	2	80			
Mean $\pm$ SD (range)	0.1 $\pm$ 0.1 (–0.3 to +0.4)	0.4 $\pm$ 0.2 (–0.3 to +1.2)	<0.001†	64.0	98.0
<b>D4</b>					
Onset difference, msec (normal $\leq$ 0.4); no. of abnormal	2	89			
Mean $\pm$ SD (range)	0.1 $\pm$ 0.2 (–0.3 to +0.6)	0.6 $\pm$ 0.3 (–0.3 to +1.8)	<0.001†	71.2	98.0
<b>D2</b>					
Digit–wrist CV, m/s (normal $\geq$ 50); no. of abnormal	3	56			
Mean $\pm$ SD (range)	56.6 $\pm$ 4.6 (45 to 68)	49.8 $\pm$ 4.7 (35 to 62)	<0.001†	44.8	97.0
Abnormal palmar, D4, or D2, n (%)	7 (7)	114 (91)	<0.001*		

\* Chi-square test.

† Mann-Whitney *U* test.

CTS = carpal tunnel syndrome; palmar = palmar median-ulnar mixed nerve comparison; D4 = digit 4 antidromic median-ulnar sensory nerve comparison; D2 = digit 2 antidromic segmental sensory nerve conduction study; CV = conduction velocity.

the NC cut-off value required to confirm MNW (defined arbitrarily as a  $\geq$ 95% PostTP) for each PreTP of MNW, and 3) the test requirements necessary to exclude MNW (defined arbitrarily as a  $\leq$ 5% PostTP) for each PreTP of MNW.

**Methods.** *Patient selection.* From January 1999 to November 2001, electrodiagnostic studies were reviewed retrospectively for all patients referred to our EMG laboratory with clinical evidence of mild CTS<sup>3,14,15</sup> or a pure musculoskeletal pain syndrome. History and physical examination were routinely performed and documented before performance of the electrophysiologic evaluation by one of the authors (H.N., D.N.H., E.L.L.).

The case definition for mild CTS was 1) nocturnal or use-related paresthesias in the median distribution with or without sensory signs on examination,<sup>3,16</sup> 2) no weakness of thumb abduction or opposition, 3) no clinical or electrophysiologic evidence of other mononeuropathies, cervical radiculopathy, or polyneuropathy, and 4) normal distal motor latency to the abductor pollicis brevis ( $\leq$ 4.4 msec).<sup>4</sup> A total of 125 consecutive hands from 112 patients met these entry criteria and composed the mild CTS patient group.

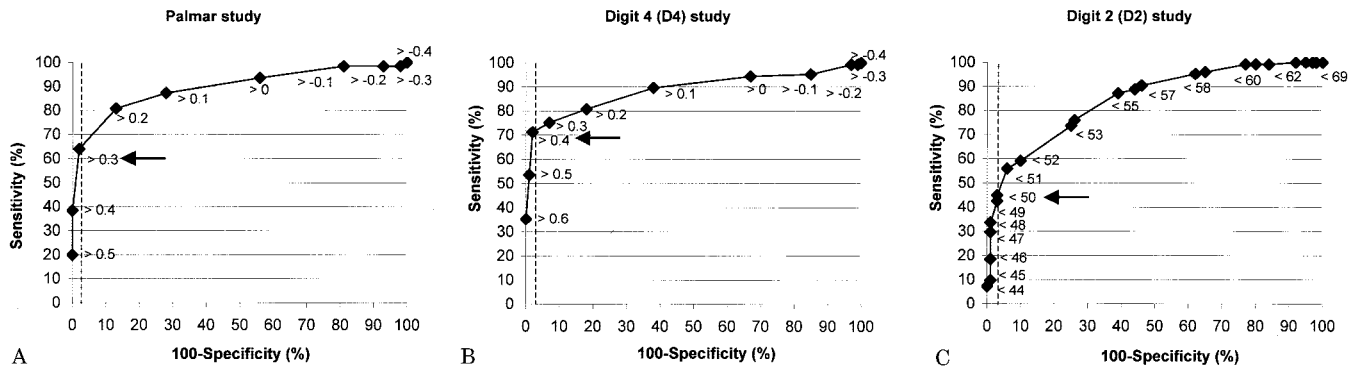
The case definition for musculoskeletal pain syndrome was 1) upper extremity pain or ache directly over joints or tendons, 2) no paresthesias, 3) no objective sensory loss, weakness of any muscles including thumb abduction and opposition, or Tinel or Phalen signs,<sup>16</sup> and 4) no clinical or electrophysiologic evidence of other mononeuropathies, cervical radiculopathy, or polyneuropathy. A total of 100 consecutive hands in 86 patients met these criteria and composed the musculoskeletal pain disease control group.

*Electrophysiologic testing.* The palmar median–ulnar mixed nerve comparison (palmar) and digit 4 antidromic median–ulnar sensory nerve comparison (D4) were performed as previously described.<sup>15,17</sup> In brief, the palmar study was performed with percutaneous stimulation of the median and ulnar nerves in the palm and surface-recording of the averaged mixed median and ulnar nerve response over the wrist using bar electrodes and a distance

of 8 cm. The onset latency of the median response was subtracted from that of the ulnar response. The D4 study was performed with percutaneous stimulation of the median and ulnar nerves at the wrist, and the averaged antidromic median and ulnar sensory response over digit 4 was recorded using ring electrodes and a distance of 13 cm. The onset latency of the median response was then subtracted from that of the ulnar response. The averaged digit 2 antidromic sensory response (D2) was recorded over the proximal phalanx of the index finger using ring electrodes after percutaneous stimulation of the median nerve at the wrist 13 cm proximally. Conduction velocity of the D2 median sensory response was then calculated using the onset latency. Finally, routine median and ulnar motor studies (standardized distal distance of 7 cm) and F responses, ulnar sensory conduction studies (standardized distance of 11 cm), and needle EMG were also performed. Hand temperature was maintained  $>$ 32 degrees centigrade.

*Data evaluation.* Receiver operator characteristic (ROC) curves.

For each of the three tests, sensitivity (the percentage of patients with MNW with a test value above [palmar or D4] or below [D2] that value) and specificity (the percentage of musculoskeletal patients with a test value at or below [palmar or D4] or at or above [D2] that value) were calculated for each possible test value. These results were then used to plot ROC curves with (100 – specificity) on the x-axis and sensitivity on the y-axis. Cut-off values for the three NC studies were defined as those that produced a specificity  $\geq$ 97% in the musculoskeletal pain control group.<sup>18</sup> Using those cut-off values, sensitivities for the three tests were then calculated and compared in the CTS group. In addition, subgroups of the patients with CTS were compared to determine whether NC test abnormalities (using cut-off test values defined above) were more frequent or more severe in patients with CTS with objective sensory loss vs patients with CTS without sensory loss, and in patients with CTS with classic sensory disturbance vs patients with CTS with probable or possible sensory disturbance (see below). The  $\chi^2$  test and Mann-Whitney *U* test were used for comparison of proportions and means with *p* < 0.05 being considered significant.



**Figure 1.** Receiver operator characteristic (ROC) curves for the (A) palmar, (B) digit 4 (D4), and (C) digit 2 (D2) studies showing the trade-off between sensitivity and specificity for various test values. Normal cut-off values (arrows) were set to obtain  $\geq 97\%$  specificity (dotted lines). For the palmar and D4 studies, the ROC curves show sharp turns at the respective cut-off values, thereby maximizing sensitivity and specificity. For the D2 study, the turn is less well defined.

**Frequency distributions.** For each of the three NC tests, the frequency distribution of individual test values was plotted for both the CTS and control groups. Interval likelihood ratios (ILR) of MNW, representing the odds that a given test value indicates MNW vs control, were then determined for each test value by the following quotient:

$$\frac{\text{Frequency of a given test value in patients with MNW}}{\text{Frequency of a given test value in patients with musculoskeletal pain}^{11,19-21}}$$

**Bayesian analysis.** Different PreTP were set arbitrarily to simulate different levels of clinical suspicion of mild CTS as follows: PreTP 0.9 (very likely), 0.75 (likely), 0.5 (equivocal), 0.25 (unlikely), and 0.1 (very unlikely).

Using the frequency distribution data, PostTP of MNW for all possible test values and PreTP were then calculated for each of the three NC tests as follows:

$$\frac{(\text{PreTP} \times \text{CTS frequency})}{(\text{PreTP} \times \text{CTS frequency}) + [(1 - \text{PreTP}) \times \text{control frequency}]^{11,19-21}}$$

(Alternatively, PostTP could also be calculated using the standard nomogram,<sup>22</sup> specifying the ILR and PreTP.) PostTP of MNW was then plotted for the entire range of possible NC test values and PreTP.

Confirmation of MNW was arbitrarily defined as a PostTP of 95% or above, and exclusion of MNW as a PostTP of 5% or less. The NC cut-off values required for confirmation of MNW were then determined and plotted for the various PreTP of disease. Conversely, the cut-off test values, ILR, and test sensitivities required to exclude MNW were also determined for the various PreTP of disease.

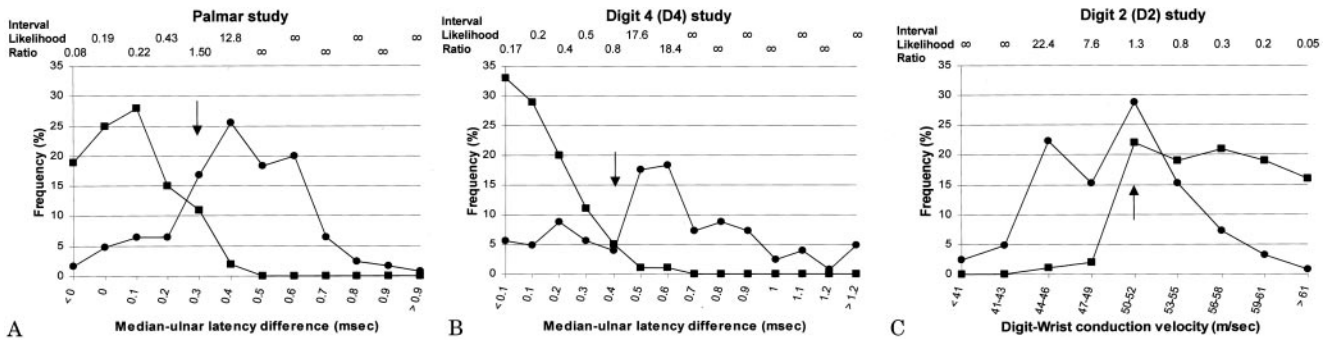
**Results. Clinical data.** In the CTS group, the distribution of median sensory symptoms (determined by questioning, not by hand diagram) was as follows: digit 1 (100 patients, 80%), digit 2 (114 patients, 91%), digit 3 (109 patients, 87%), and digit 4 (76 patients, 61%). Overall, the sensory distribution was classified as per Katz et al.<sup>23,24</sup> as classic (symptoms affect  $\geq 2$  of digits 1 to 3) in 47 hands (38%), probable (classic + symptoms in the palm) in 59 (47%), and possible (symptoms involve only 1 of digits 1 to 3) in 19 (15%). Patients with additional minor sensory symptoms in the fifth digit or proximal to the wrist were not excluded.<sup>25</sup> Loss of sensation to touch–pressure or reduced sharp–dull discrimination in the median nerve distribution was found in 47 hands (38%). Tinel sign over the median nerve at the wrist and Phalen sign were positive in 37 (30%) and 27 (22%).<sup>16</sup> In the musculoskeletal pain disease control group, 45 (45%) had focal tenderness on examination over tendons or joints that reproduced their symptom. Age and sex were similar between the CTS and disease control groups (table,  $p = \text{NS}$ ).

**ROC curves.** ROC curves for each of the three tests and their respective normal–abnormal cut-off values are shown in figure 1. Palmar and D4 curves show sharper turns than does the D2 curve for the cut-off values that optimize specificity at 97%.<sup>13,18,26</sup> To achieve this specificity, the following cut-off values were required:  $>0.3$  msec (palmar) and  $>0.4$  msec (D4) of the median over the respective ulnar onset latencies, and digit to wrist conduction velocity (CV)  $<50$  m/sec (D2). Using these cut-off values, the sensitivity for the palmar (64%) and D4 (71%) median–ulnar comparison studies was not different ( $p = \text{NS}$ ) but the sensitivity for both the palmar and D4 studies was greater than that of the D2 study (45%,  $p < 0.01$ ). Combining the results of all three tests, NC abnormalities were identified in 114 patients with CTS (91%) but only in 7 controls (7%) (see the table,  $p < 0.001$ ).

Of note, the cut-offs and sensitivities found above were exactly the same when all the left-hand studies were eliminated in those patients with bilateral studies, thus reducing the MNW group to 112 patients and the musculoskeletal pain group to 86 patients. Hence, inclusion of patients with bilateral studies did not appear to bias the results. Second, as expected, digit 2 conduction velocity was negatively correlated with age in both the musculoskeletal pain ( $r = -0.32$ ) and MNW ( $r = -0.40$ ) groups, but there was no correlation with age for the D4 or palmar studies. Third, prevalence and severity of electrodiagnostic test abnormalities were not different in the CTS patient subgroups with objective sensory loss vs the CTS subgroup without sensory loss, or in the CTS subgroups with a classic hand diagram vs the subgroup with a probable or the subgroup with a possible hand diagram. Finally, there was no correlation between the presence (or absence) of symptoms in a given digit and abnormalities (or lack thereof) on one of the three NC tests; that is, symptom location did not predict electrophysiologic test results.

**Frequency distributions.** The frequency of NC test values and the resultant interval ILR of MNW for all test values are shown in figure 2 for the CTS and control groups. It can be seen that there is less overlap of test values in the palmar and D4 distributions than for the D2 distribution. In addition, this kind of display is useful in the electrodiagnostic laboratory as it shows the relative likelihood of CTS vs control for a given test interval (e.g., ILR). By contrast, the ROC curve alone can provide similar data for all test values greater than a given cut-off value, but not for a given test value or small interval of test values.

**Bayesian analysis.** PostTP of MNW is shown for the three tests in figure 3, and examples of the calculation of PostTP are shown in the Appendix. With the cut-off values described above, PostTP is higher for palmar and D4 than for D2 studies, and depends on both the actual test value and the PreTP. For all three tests, with high PreTP, a mildly abnormal test value (e.g., digit 4 = 0.5 msec) results in a very high PostTP (i.e.,  $\geq 95\%$ ), whereas with low PreTP, the same value results in only an intermediate PostTP. By contrast, a very abnormal value of 0.7 msec for the same test results in a PostTP of 100% regardless of PreTP. Conversely, there were no normal palmar or D4 NC test values that could exclude MNW (e.g., achieve a PostTP  $\leq 5\%$ ) when PreTP was higher than 25%. Only extremely high D2 velocities of  $>61$



**Figure 2.** Frequency distribution of electrodiagnostic test values in the carpal tunnel syndrome (CTS) and control groups is plotted and the interval likelihood ratio of median neuropathy at the wrist is displayed above each test value or interval. Arrows indicate upper limits of normal as determined above. Note the sharp increase in frequency in the CTS group for the (A) palmar (above 0.3 msec) and (B) digit 4 (D4) (above 0.4 msec) median-ulnar comparison studies, and increase in frequency in the control group below 0.3 msec (palmar) and below 0.4 msec (D4). In contrast, there is broader overlap in frequencies of (C) D2 conduction velocities in the CTS and control groups. ● = CTS; ■ = control.

m/sec could exclude CTS for PreTP of disease of 50% or below, but could not exclude CTS for PreTP of 75 or 90%.

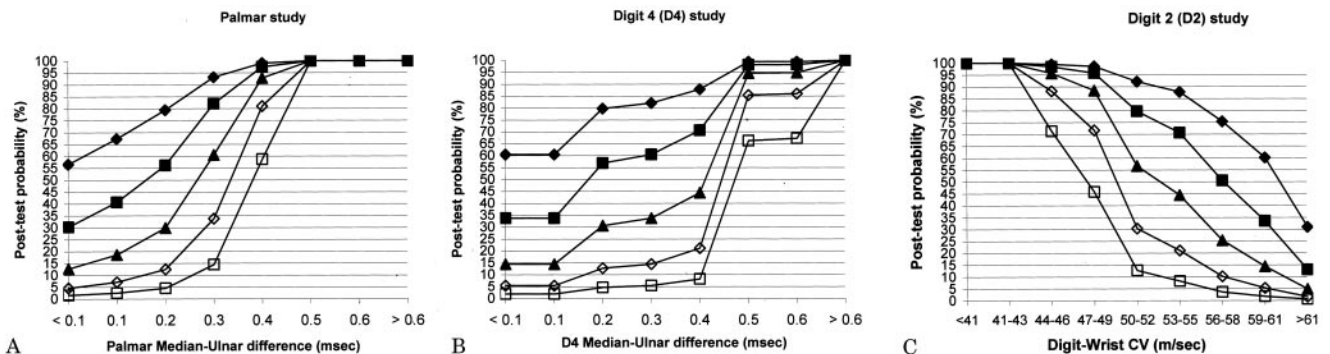
The cut-off test value required to confirm MNW and achieve a 95% PostTP varies according to PreTP of disease (figure 4). When PreTP of MNW is low, a 95% PostTP of MNW can only be attained when the cut-off values are adjusted upwards to 0.7 msec for D4 and 0.5 msec for palmar (see figure 4A) and downwards to 44 m/sec for D2 (see figure 4B).

In this study, NC tests could not exclude MNW in patients with symptoms of mild CTS and high PreTP, because for all NC test values within the normal range, the ILR are not low enough (or their sensitivities are not high enough). For example, to achieve a 5% or lower PostTP of MNW, very low ILR of 0.05 or less (figure 5A) or very high sensitivities of about 95% or more (see figure 5B) are required for PreTP of 50% or greater. Even when all three NC tests are combined, yielding a combined sensitivity of 91% and a combined specificity of 93%, MNW cannot be excluded when PreTP is 50% or above (see figure 5C).

**Discussion.** This study is the first Bayesian analysis in the English literature that examines the diagnostic power of NC tests commonly used to confirm MNW in patients with CTS, and one of only a few electrophysiologic studies that utilizes ROC curves,<sup>26-30</sup> frequency distributions, or ILR. The only other Bayesian

analysis of electrophysiologic testing for CTS was that of Madrazo et al.,<sup>31</sup> who examined the effect of PreTP on efficacy of electrodiagnostic testing in general, but did not determine the PostTP of MNW for individual NC test values and the entire range of PreTP, as in this study.

Rather than viewing a NC test result as normal or abnormal, our study looks at NC testing as an instrument to modify PreTP of MNW. The resultant PostTP of MNW depends on estimates of PreTP, on the actual NC test measurement itself, and on knowledge of the test's frequency distribution (e.g., providing frequencies for all NC test outcomes in the CTS and control group), ILR, and Bayes theorem. This analysis shows that to confirm MNW, the electrophysiologist cannot apply the same cut-off test value to patients with different PreTP of disease, because a patient with low PreTP requires a higher D4 or palmar or lower D2 velocity cut-off to achieve a 95% PostTP of MNW than does a patient with a high PreTP. Conversely, to exclude MNW, very high test



**Figure 3.** Post-test probabilities (PostTP) of median neuropathy at the wrist (MNW) calculated with different pretest probabilities (PreTP) of MNW for each test value obtained. Note: 1) PostTP depends on both the actual test value and the PreTP, with higher PreTP yielding higher PostTP; 2) PostTP is always higher for palmar (A) and digit 4 (D4) (B) studies than for the digit 2 (D2) (C) study for the standard cut-off values (see the table); 3) a borderline abnormal test value (e.g., palmar = 0.4 msec, D4 = 0.5 msec, D2 = 47 to 49 m/sec) yields very high PostTP ( $\geq 95\%$ ) when PreTP is high, whereas the same test values result in only intermediate PostTP when PreTP is low; and 4) by contrast, abnormal test values (e.g., 0.7 msec in palmar and D4 and 43 m/sec in D2) yield a PostTP of 100% regardless of the PreTP. ◆ = PreTP 0.9; ■ = 0.75; ▲ = 0.5; ◇ = 0.25; □ = 0.1.

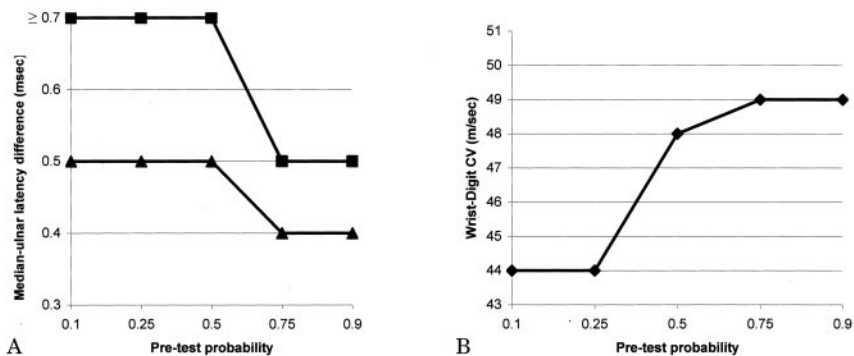


Figure 4. (A) Digit 4 [D4] and palmar and (B) digit 2 [D2] show the test cut-offs required to achieve a post-test probability (PostTP) of 95% or greater for various pre-test probabilities (PreTP) of median neuropathy at the wrist (MNW). Note that higher palmar and D4 median-ulnar comparison and lower D2 conduction velocity cut-offs are necessary to confirm MNW when PreTP of MNW declines. ■ = D4; ▲ = palmar; ◆ = D2.

sensitivities of 95 to 100% are required when PreTP is above 50%, well above those of these three tests, alone or combined. Finally, this study also confirmed that the palmar and D4 studies have comparable sensitivities in mild CTS, ROC curve analyses, and impact on PostTP, whereas the D2 study is a less powerful diagnostic test.

ROC curves have been underutilized in the evaluation of electrodiagnostic tests in general. In the CTS literature, there are only five studies employing ROC curves.<sup>26-30</sup> Their inherent advantage is that they display the sensitivity and specificity of all possible cut-off values for a given test. From the ROC curve one can easily 1) visualize the trade-off between sensitivity and specificity for a given test value—the higher the one, the lower the other; 2) determine an optimal cut-off value in the traditional sense<sup>28-30</sup>; 3) determine the cut-off to achieve a given specificity; 4) given a set specificity, compare the sensitivities of different tests; and 5) evaluate the diagnostic power of the test by calculating the area under the ROC curve.<sup>27,28</sup> In this study, nearly all of these ROC applications were achieved. When setting specificity at 97%, the optimal cut-offs for the various tests essentially matched those reported by others (see above),

with the sensitivities of digit 4 and palmar studies being superior to that of the digit 2 study.

Although useful for comparison of diagnostic tests, ROC curves do not provide data on a given test value or small interval of test values, but rather yield data for all values above (or below) a given test cut-off value. Therefore, they are less useful in clinical decision making than frequency distributions and ILR. For example, if the electrophysiologist wants to know the likelihood of MNW vs control for a patient with a D2 conduction velocity of 48 m/sec (e.g., within the interval from 47 to 49 m/sec), the ROC curve alone does not suffice for this purpose. By contrast, the frequency distributions do allow easy determination of ILR for MNW vs control for given test value intervals (see figure 2) by simply dividing the CTS frequency by the control frequency for each interval. Thus, for a patient with a D2 velocity of 48 m/sec, the likelihood of MNW is 7.6 times that of control. However, the remaining drawback of the frequency distributions and ILR is that they do not account for PreTP of disease.

Bayes theorem does incorporate PreTP of disease, and with the frequency distribution data or ILR allows for calculation of PostTP of MNW for the entire

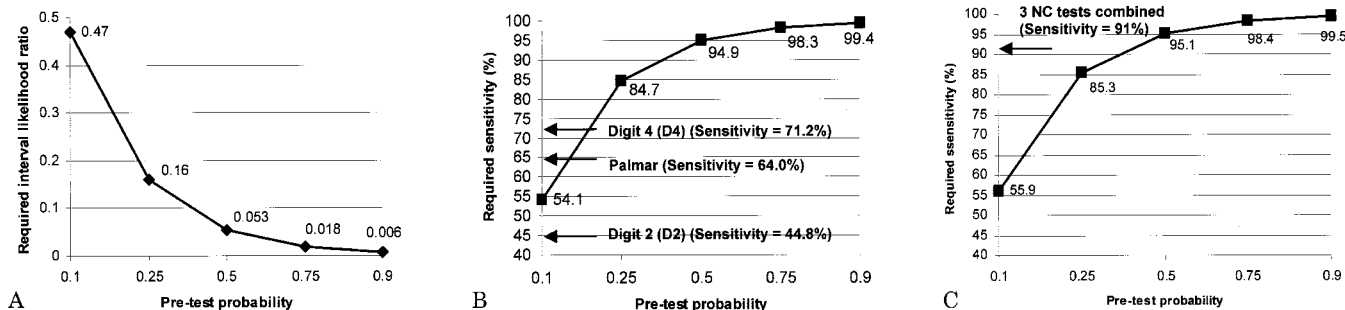


Figure 5. (A) Single nerve conduction (NC) test: Interval likelihood ratios (ILR) required to attain a post-test probability (PostTP) of 0.05 or less for the range of pretest probabilities (PreTP). Note that for PreTP above 50%, the only NC test value (see figure 2) with an ILR low enough (e.g., <0.053) to attain a PostTP of <0.05 is a digit 2 (D2) conduction velocity of >61 m/sec. (B) Single NC test: Sensitivities required to attain a PostTP of 0.05 or less for the range of PreTP, assuming a test specificity of 97% and a negative single test outcome. Arrows indicate sensitivities of the three NC tests. (C) Three NC tests combined: Test sensitivities required to attain a PostTP of 0.05 or less for the range of PreTP, assuming a combined test specificity of 93% and a negative NC test outcome on all three tests. Arrow indicates 91% sensitivity of the 3 NC tests combined (specificity = 93%). Note that single NC tests (A) have a maximum sensitivity of 71% and therefore can exclude median neuropathy at the wrist (MNW) for a PreTP of 10%, but not for 25% or above. Even when all three tests are combined (B), with a sensitivity of 91%, MNW can be excluded for PreTP up to 25%, but not for 50% or above.

range of individual test values, and a given a priori PreTP. The advantage of Bayesian analysis is that it 1) maximizes the diagnostic information of a given test outcome, rather than simply grouping test outcomes as normal and abnormal, and 2) accounts for PreTP of disease. In the example of a patient with a D2 velocity of 48 m/sec, and a likelihood ratio of 7.6, the PostTP of MNW still varies dramatically from about 45% to 98%, depending on PreTP (see figure 3, Appendix).

PreTP of disease is a foreign concept to most neurologists and electrophysiologists, and more importantly, at least for mild MNW, it is not easily quantifiable at this time. However, it is widely accepted that electrodiagnostic testing is an extension of the clinical evaluation, and that electrophysiologic results must be interpreted in light of the clinical picture. Most clinical electrophysiologists are wary of electrodiagnosis of any disorder based on mildly abnormal electrical findings in the absence of compatible clinical findings and therefore this study quantitates what most good clinicians do anyway. This is particularly true for technically challenging cases in which test reproducibility (e.g., test–retest variability) may be suboptimal.

The major question raised by our study, and one that requires further research, is whether PreTP of mild MNW can be quantitated using clinical and demographic characteristics. Others have attempted this in various non-neurologic disorders using clinical features.<sup>32–35</sup> We are aware of four studies that have attempted this in CTS,<sup>23–25,36</sup> using the hand diagram,<sup>23–25,36</sup> a hand sensory questionnaire,<sup>25</sup> and Phalen and Tinel signs,<sup>24,36</sup> but the authors used electrodiagnostic NC test values as the gold standard for CTS (see below). Despite this drawback, using these and other clinical features,<sup>16</sup> we believe it is possible to decide whether the diagnosis of mild MNW is at least unlikely, equivocal, or likely and therefore could attempt to determine an approximate PreTP as 25%, 50%, or 75%. Whether estimates of PreTP can be fine-tuned further to add in two other PreTP (10%, very unlikely, and 90%, very likely) as in this study is uncertain, but there is little doubt about the basic message of this analysis; namely, that borderline abnormal results (e.g., D4 = 0.5 msec) cannot confirm MNW (e.g., achieve a PostTP of 95%) in the absence of clinical features strongly compatible with the diagnosis, whereas very abnormal results (e.g., D4 = 0.7 msec) can. Conversely, there are virtually no normal NC test values with low enough ILR to exclude MNW when PreTP of disease is 50% or above.

This study has limitations, the most important being lack of a gold standard for diagnosis of mild CTS. We used clinical criteria for this purpose, as recommended by the American Association of Electrodiagnostic Medicine<sup>4</sup> and the American Public Health Association<sup>3</sup> for studies of CTS. We were particularly interested in patients with mild CTS, as these are the most challenging for the electrophysiologist and require the most sensitive tests (e.g., pal-

mar and digit 4).<sup>4</sup> The disease control group was chosen deliberately because electrophysiologists are not referred normal subjects, and it is particularly important to be able to distinguish patients with mild CTS from those with musculoskeletal pain. Given the limitation of a clinical classification system, it is striking that 1) the standard cut-off values obtained for the palmar (0.3 msec) and digit 4 (0.4 msec) studies in our patients are exactly the same as those reported by others,<sup>4,8,15,30</sup> 2) the sensitivity of the palmar and digit 4 in our study was 64 to 71%—very similar to that reported by others in mild CTS,<sup>14,17,30</sup> and 3) the combined sensitivity of all three tests was 91% with a 93% specificity. Still, the lack of a gold standard may have resulted in higher false positives and false negatives and increased the overlap of the NC test value frequency distribution curves shown in figure 2. Nevertheless, it is unlikely that cut-off values with a 100% sensitivity and specificity exist for these NC tests, and therefore Bayesian analysis is likely to be informative even if there was a true gold standard.

Test- and diagnostic-review bias<sup>37</sup> could have led to inaccurate group classification of CTS and disease control patients in our study. However, because the Nicolet microprocessor automatically marks the NC test latencies, test-review bias is probably small. Diagnostic-review bias is possible in this study because the clinicians were not blinded to the test results, but is also likely small because the documented clinical evaluation always preceded the NC testing.

There are two points about likelihood ratios that require comment. First, this analysis assumes that ILR do not change significantly as PreTP of MNW varies. Although ILR are thought to be stable with varying PreTP,<sup>11</sup> it is conceivable that as PreTP declines, the frequency distribution curve for the CTS group could approach that of the control group, thereby shifting the ILR toward 1.0. But, to the extent this occurs, PostTP shown in figure 3 for abnormal NC test values would decline and those for normal values would rise, thus requiring even more conservative cut-off NC test values to confirm MNW (e.g., higher palmar and D4 values) and also making it more difficult to exclude MNW. Second, the calculated ILR are average ILR for the heterogeneous CTS group. Although there were no electrodiagnostic differences between CTS subgroups, it is likely that ILR for a given test interval may vary according to disease characteristics, age, or sex.<sup>38</sup> This is clearly the case for the digit 2 study because D2 velocity was found to vary inversely with age, and therefore a given low D2 velocity in a 20-year-old actually has a higher ILR than does the same velocity in a 60-year-old patient. This is probably true for other anthropometric measurements,<sup>39,40</sup> and therefore, the analysis in this study might be further refined by the use of percentiles and normal deviates rather than raw NC values.<sup>41</sup>

Finally, it could be argued that our arbitrary requirement of a 95% PostTP to confirm MNW is too restrictive. Our reason for setting such a stringent

rule-in criterion was to minimize the consequences of a false positive diagnosis (e.g., unnecessary surgical decompression), and to set stretch targets for optimal clinical decision-making. Despite its limitations, this study underscores the importance of considering PreTP and the actual NC test value when attempting to confirm or exclude MNW. It also emphasizes the utility of the ILR over that of more traditional measures when presenting NC test data, and suggests that this kind of analysis would be informative for other neurodiagnostic applications. Ultimately, however, methods to quantitate PreTP will be required to maximize the diagnostic potential of this approach.

## Acknowledgment

The senior author (E.L.L.) thanks William C. Taylor, MD, for his memorable, one-on-one Bayes theorem teaching sessions over 20 years ago. The authors also acknowledge the technical expertise of Marlene Hogan and Michele Ferguson, the editorial expertise of Laura Aronberg, and the database development by James Forrester.

## Appendix

Calculation of post-test probability (PostTP) for three different pretest probabilities (PreTP) (0.25, 0.50, 0.75) assuming digit 2 conduction velocity = 48 m/sec, and therefore, per figure 2, carpal tunnel syndrome frequency = 15.2% and control frequency = 2%

1) For PreTP = 0.25:

$$\text{PostTP} = \frac{0.25 \times 15.2}{(0.25 \times 15.2) + [(1 - 0.25) \times 2]} = 0.717 \text{ (71.7\%)}$$

2) For PreTP = 0.50:

$$\text{PostTP} = \frac{0.50 \times 15.2}{(0.50 \times 15.2) + [(1 - 0.50) \times 2]} = 0.884 \text{ (88.4\%)}$$

3) For PreTP = 0.75:

$$\text{PostTP} = \frac{0.75 \times 15.2}{(0.75 \times 15.2) + [(1 - 0.75) \times 2]} = 0.958 \text{ (95.8\%)}$$

## References

- Phalen GS. The carpal tunnel syndrome. *J Bone Joint Surg* 1966;48A:211–228.
- Gelberman RH, Rydevik BL, Pess GM, Szabo RM, Lundborg G. Carpal tunnel syndrome. *Orthop Clin North Am* 1988;19:115–124.
- Rempel D, Evanoff B, Amadio PC, et al. Consensus criteria for the classification of carpal tunnel syndrome in epidemiologic studies. *Am J Public Health* 1998;88:1447–1451.
- Jablecki CK, Andary MT, So YT, Wilkins DE, Williams FH. Literature reviews of the usefulness of nerve conduction studies and electromyography for the evaluation of patients with carpal tunnel syndrome. *Muscle Nerve* 1993;16:1392–1414.
- Buchthal F, Rosenfalck A. Sensory conduction from digit to palm and from palm to wrist in the carpal tunnel syndrome. *J Neurol Neurosurg Psychiatry* 1971;34:243–252.
- Kimura J. The carpal tunnel syndrome: localization of conduction abnormalities within the distal segment of the median nerve. *Brain* 1979;102:619–635.
- Daube JR. Percutaneous palmar median nerve stimulation for carpal tunnel syndrome. *Electroencephalogr Clin Neurophysiol* 1977;43:139–140.
- Stevens JC. AAEM minimonograph #26: the electrodiagnosis of carpal tunnel syndrome. *Muscle Nerve* 1997;20:1477–1486.
- Johnson EW, Kukla RD, Wongsam E, Piedmont A. Sensory latencies to the ring finger: normal values and relation to carpal tunnel syndrome. *Arch Phys Med Rehabil* 1981;62:206–208.
- Mills KR. Orthodromic sensory action potentials from palmar stimulation in diagnosis of carpal tunnel syndrome. *J Neurol Neurosurg Psychiatry* 1985;48:250–255.
- Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine. How to practice and teach EBM, 2nd ed. Edinburgh: Churchill Livingstone, 2000.
- Sackett DL, Straus S. On some clinically useful measures of the accuracy of diagnostic tests. *Evidence-Based Medicine* 1998;3:68–70.

- Schulzer M. Diagnostic tests: a statistical review. *Muscle Nerve* 1994;17:815–819.
- Girlanda P, Quartarone A, Sinicropi S, et al. Electrophysiological studies in mild idiopathic carpal tunnel syndrome. *Electroencephalogr Clin Neurophysiol* 1998;109:44–49.
- Preston DC, Ross MH, Kothari MJ, Plotkin GM, Venkatesh S, Logigian EL. The median-ulnar latency difference studies are comparable in mild carpal tunnel syndrome. *Muscle Nerve* 1994;17:1469–1471.
- D'Arcy CA, McGee S. Does this patient have carpal tunnel syndrome? [The rational clinical examination.] *JAMA* 2000;283:3110–3117.
- Uncini A, Di Muzio A, Awad J, Manente G, Tafuro M, Gambi D. Sensitivity of three median-to-ulnar comparative tests in diagnosis of mild carpal tunnel syndrome. *Muscle Nerve* 1993;16:1366–1373.
- Rivner MH. Statistical errors and their effect on electrodiagnostic medicine. *Muscle Nerve* 1994;17:811–814.
- Schmitz N, Kruse J, Tress W. Application of stratum-specific likelihood ratios in mental health screening. *Soc Psychiatry Psychiatr Epidemiol* 2000;35:375–379.
- Furukawa T, Hirai T, Kitamura T, Takahashi K. Application of the Center for Epidemiologic Studies Depression Scale among first-visit psychiatric patients: a new approach to improve its performance. *J Affect Disord* 1997;46:1–13.
- Sonis J. How to use and interpret interval likelihood ratios. *Fam Med* 1999;31:432–437.
- Fagan TJ. Nomogram for Bayes's theorem. *N Engl J Med* 1975;293:257.
- Katz JN, Stirrat C, Larson MG, Fossel AH, Eaton HM, Liang MH. A self-administered hand symptom diagram in the diagnosis and epidemiologic study of carpal tunnel syndrome. *J Rheumatol* 1990;17:1495–1498.
- Katz JN, Larson MG, Sabra A, et al. Carpal tunnel syndrome: diagnostic utility of history and physical examination findings. *Ann Intern Med* 1990;112:321–327.
- Stevens JC, Smith BE, Weaver AL, Bosch EP, Deen HG Jr, Wilkens JA. Symptoms of 100 patients with electromyographically verified carpal tunnel syndrome. *Muscle Nerve* 1999;22:1448–1456.
- Eisen A, Schulzer M, Pant B, et al. Receiver operating characteristic curve analysis in the prediction of carpal tunnel syndrome: a model for reporting electrophysiological data. *Muscle Nerve* 1993;16:787–796.
- Leffler CT, Gozani SN, Cros D. Median neuropathy at the wrist: diagnostic utility of clinical findings and an automated electrodiagnostic device. *J Occup Environ Med* 2000;42:398–409.
- Atroshi I, Johnsson R. Evaluation of portable nerve conduction testing in the diagnosis of carpal tunnel syndrome. *J Hand Surg* 1996;21A:651–654.
- Gunnarsson LG, Amilon A, Hellstrand P, Leissner P, Philipson L. The diagnosis of carpal tunnel syndrome: sensitivity and specificity of some clinical and electrophysiological tests. *J Hand Surg* 1997;22B:34–37.
- Robinson LR, Mickesen PJ, Wang L. Strategies for analyzing nerve conduction data: superiority of a summary index over single tests. *Muscle Nerve* 1998;21:1166–1171.
- Madrazo J, Marin I, Bringas A, Fernandez A. The efficacy of neurophysiological examination in the diagnosis of carpal tunnel syndrome. *Revista de Neurologia* 2000;30:1005–1008.
- Wells PS, Hirsh J, Anderson DR, et al. Accuracy of clinical assessment of deep-vein thrombosis. *Lancet* 1995;345:1326–1330.
- Gibbons RJ, Balady GJ, Beasley JW, et al. ACC/AHA guidelines for exercise testing. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Exercise Testing). *J Am Coll Cardiol* 1997;30:260–311.
- Welch RD, Zelenski RJ, Shamsa F, et al. Pretest probability assessment for selective rest sestamibi scans in stable chest pain patients. *Am J Emerg Med* 2000;18:789–792.
- Morise AP. Comparison of the Diamond-Forrester method and a new score to estimate the pretest probability of coronary disease before exercise testing. *Am Heart J* 1999;138:740–745.
- O'Gradaigh D, Merry P. A diagnostic algorithm for carpal tunnel syndrome based on Bayes's theorem. *Rheumatology* 2000;39:1040–1041.
- Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926–930.
- Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and Bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology* 1997;8:12–17.
- Stetson DS, Albers JW, Silverstein BA, Wolfe RA. Effects of age, sex and anthropometric factors on nerve conduction measures. *Muscle Nerve* 1992;15:1095–1104.
- Salerno DF, Franzblau A, Werner RA, Bromberg MB, Armstrong TJ, Albers JW. Median and ulnar nerve conduction studies among workers: normative values. *Muscle Nerve* 1998;21:999–1005.
- Dyck PJ, O'Brien PC, Litchy WJ, Harper CM, Daube JR, Dyck PJB. Use of percentiles and normal deviates to express nerve conduction and other test abnormalities. *Muscle Nerve* 2001;24:307–310.