From previous studies, abnormality of nerve conduction (NC) is known to be a sensitive, objective, and quantitative indication typical of diabetic sensorimotor polynuropathy (DSPN).1–5 To be such an indication, the NC procedures must be done proficiently, attributes must be compared with reference values corrected for applicable variables (e.g., age, gender, height, and weight), criteria used for diagnosis must be exactly defined (prespecified) and representative of neurophysiological abnormalities found in DSPN, and they should be shown to be both sensitive and specific for the diagnosis of DSPN based on study of healthy subject and diabetes mellitus (DM) cohorts. In this study we tested eight specific criteria for the NC diagnosis of DSPN.

The performance of eight NC criteria was assessed in previously studied cohort databases of population-representative healthy subjects [i.e., the Rochester Diabetic Neuropathy Study—healthy subjects (RDNS-HS)] and DM (RDNS).6–8

The first group of criteria (1–4) used defined percentile (≤1st/≥99th, ≤2.5th/≥97.5th, or ≤5th/≥99th) threshold abnormalities of individual or multiple attributes of NC in single or multiple nerves for diagnosis. The second group (criteria 5–8) used summated normal deviate (from percentiles) scores with the percentile value of the composite score based on the RDNS-HS study cohort.

The different criteria were compared by the degree of exact prespecification of abnormalities assessed, whether attribute abnormalities are representative of neurophysiological abnormalities in DSPN, and by the degree of specificity obtained in study of a healthy subject cohort and sensitivity in a DM cohort.9

METHODS

The eight diagnostic NC criteria for DSPN were compared in databases of a population-based healthy subject cohort (RDNS-HS)6,10 and a DM cohort (RDNS),7 both at first evaluation.

HS and DM Cohorts and Description of Availability of NC Data. This study was possible because we had previously made measurements of 14 NC attributes in population-based studies of a cohort of individuals without neurological disease or medical conditions predisposing them to polyneuropathy in Olmsted County, Minnesota, USA (the RDNS-HS cohort)8,10 and of persons with DM in the same community (RDNS).7 The protocol for the study was approved by the institutional review board of the Mayo Clinic. The methods of testing attributes of NC are standard ones employed at the Mayo Clinic and have been extensively described previously.7,8,10 The RDNS-HS cohort values were used to set percentiles specific for the applicable variables of age, gender, height, and weight.8,10 The NC values at prevalence of a population-representative group of patients with DM and as corrected for applicable variables has also been extensively described.6,7

We used data from these two cohorts, making the assumption that the RDNS-HS cohort should have abnormality in only 5%, 2.5%, or 1%—the prespecified criteria for abnormality. Using these...
specificity levels in the DM cohort (RDNS), we tested for percent sensitivity, and a large difference of NC abnormality was expected between RDNS-HS and RDNS.

**NC Criteria for Distal Symmetric Length-Dependent Sensorimotor Polyneuropathy.** NC criteria for typical DPN (distal symmetric length-dependent sensorimotor polyneuropathy, DSPN) are listed in Table 1, but are described here in greater detail. Criterion 1 is abnormality (by a prespecified percentile value) of any one or more attribute of NC of any anatomical nerve tested. Criterion 2 (used in early RDNS cohort studies) is abnormality of any anatomical nerve tested. Criterion 2 (used in early RDNS cohort studies) is abnormality of any anatomical nerve tested. Criterion 3 is abnormality of the composite score set at a defined percentile level.

**Description of RDNS-HS and DM Cohorts.** Population-based cohort studies of the Olmsted County, Minnesota, USA, population are possible because essentially all patients in this county receive their medical care at either the Mayo Clinic or the Olmsted Medical Center, which share a common registry of diagnoses and laboratory test results; approximately 70% of inhabitants are medically

A brief description of the derivation of composite scores 5–8 follows. First, the percentile values of the studied attributes are corrected for applicable variables of age, gender, height, weight, or other anthropomorphic variables. Percentile abnormality of all attributes used in the composite score are expressed in the same tail of the normal distribution (e.g., in our criteria 5–8) in the upper tail, but the lower tail would also be suitable. This is done by simple conversion of lower to upper tail values (e.g., ≤1st to ≥99th or 45th–55th percentiles). All percentile values are expressed as normal deviate values from percentiles. The mean normal deviate value of the measurable attributes is multiplied by the number of attributes of the composite score (i.e., ×2 in criteria 5 and 6, ×4 in criterion 7, and ×6 in criterion 8). This mathematical manipulation is needed when some attributes may be unmeasurable, such as when amplitude is 0 (a measurable value), and conduction velocity and distal latency usually cannot be estimated. The value of the derived composite normal deviate from percentiles (nd) score is then determined from estimated percentiles of the composite score assessed in a healthy subject cohort (in this case, the RDNS-HS).

The 14 attributes of NC measured included: amplitude (CMAP), velocity (MNCV), and distal latency motor nerve distal latency (MNDL) of motor fibers of the ulnar, fibular and tibial nerves; F latency of the ulnar, fibular, and tibial nerves; and amplitude of the ulnar and sural nerves (SNAP).

**Description of RDNS-HS and DM Cohorts.** Population-based cohort studies of the Olmsted County, Minnesota, USA, population are possible because essentially all patients in this county receive their medical care at either the Mayo Clinic or the Olmsted Medical Center, which share a common registry of diagnoses and laboratory test results; approximately 70% of inhabitants are medically

### Table 1. Nerve conduction abnormality in the RDNS and RDNS-HS cohorts using different criteria.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>RDNS-HS* (N = 330): n (%) abnormal</th>
<th>RDNS (N = 456): prevalence at first visit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5th/95th</td>
<td>2.5th/97.5th</td>
</tr>
<tr>
<td>Criterion 1: &gt;1 of 12 NC attributes abnormal†</td>
<td>123 (37.3)</td>
<td>57 (17.3)</td>
</tr>
<tr>
<td>Criterion 2: &gt;1 abnormal in 2 separate nerves</td>
<td>37 (11.2)</td>
<td>8 (2.4)</td>
</tr>
<tr>
<td>Criterion 3: &gt;1 abnormal in 2 separate nerves (1 is sural)</td>
<td>24 (7.3)</td>
<td>7 (2.1)</td>
</tr>
<tr>
<td>Criterion 4: Fibular CV abnormal and sural amplitude abnormal</td>
<td>2 (0.6)</td>
<td>2 (0.6)</td>
</tr>
<tr>
<td>Criterion 5: ≥2 NC nds abnormal (fibular CV and sural amplitude)</td>
<td>17 (5.2)</td>
<td>9 (2.7)</td>
</tr>
<tr>
<td>Criterion 6: ≥2 NC nds abnormal (fibular CV and tibial CV)</td>
<td>17 (5.2)</td>
<td>9 (2.7)</td>
</tr>
<tr>
<td>Criterion 7: ≥5 NC nds abnormal</td>
<td>17 (5.2)</td>
<td>9 (2.7)</td>
</tr>
<tr>
<td>Criterion 8: ≥6 NC nds abnormal</td>
<td>17 (5.2)</td>
<td>9 (2.7)</td>
</tr>
</tbody>
</table>

*Based on theoretical considerations (Bonferroni modeling) the following abnormal frequencies would be expected based on type 1 error and lack of linkage among attributes studies.

†Of 330 x 12 = 3960 nerve attributes tested, 197 (5.0%) are abnormal at the 95th, 75 (1.9%) at the 97.5th, and 36 (0.9%) at the 99th.
In decreasing order, the following frequencies of scores of defined composite attribute scores (the corrected for applicable variables and as sum prevalence were used in this analysis. The attrib-
utes of NC at cohort of volunteers with DM (initially by NDDG and later by ADA criteria). Attributes of NC at prevalence were used in this analysis. The attrib-
utes of NC were expressed as ns from percentiles corrected for applicable variables and as sum scores of defined composite attribute scores (the same approach as used in the RDNS-HS).

**RESULTS**

**Frequency of Individual Attributes of NC Abnormality in RDNS Cohort and other Performance Characteristics.** In decreasing order, the following frequencies of individual attributes of NC abnormality at ≤2.5th/97.5th percentiles were obtained in the RDNS cohort at prevalence: fibular MNCV (26.3%); sural SNAP (25.4%); tibial MNCV (24.8%); ulnar MNCV (21.3%); tibial F latency (16.9%); ulnar F latency (16.0%); and then, in lesser frequencies, the other attributes. Among these six attributes, abnormality of a single NC attribute only in DM patients was observed in the following numbers and percentages: fibular MNCV 14 (3.1%); sural SNAP 22 (4.8%); tibial MNCV 17 (3.7%); ulnar MNCV 11 (2.4%); tibial F-wave latency 6 (1.3%); and ulnar F-wave latency 7 (1.5%). However, none of these patients with abnormality of only one attribute was judged to have a clinical mononeuropathy of that nerve.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>5th/95th percentile</th>
<th>2.5th/97.5th percentile</th>
<th>1st/99th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kappa</td>
<td>P&lt;0.0001</td>
<td>Kappa</td>
<td>P&lt;0.0001</td>
</tr>
<tr>
<td>Criterion 3: &gt;1 abnormal in 2 separate nerves (1 sural)</td>
<td>0.56</td>
<td>&lt;0.0001</td>
<td>0.49</td>
</tr>
<tr>
<td>Criterion 4: Fibular CV abnormal and sural amplitude abnormal</td>
<td>0.20</td>
<td>&lt;0.0001</td>
<td>0.36</td>
</tr>
<tr>
<td>Criterion 5: Σ 2 NC nds abnormal</td>
<td>0.32</td>
<td>&lt;0.0001</td>
<td>0.09</td>
</tr>
<tr>
<td>Criterion 6: Σ 6 NC nds abnormal</td>
<td>0.57</td>
<td>&lt;0.0001</td>
<td>0.43</td>
</tr>
</tbody>
</table>

**Agreement between Pairs of NC Attributes for Diagnosis of Polyneuropathy.** We tested for agreement between pairs of NC attributes for the diagnosis of DSPN using the kappa coefficient (Table 2). Note that there was a highly significant association for all but one pair of NC attributes tested. For many attribute pairs, there was good agreement for the diagnosis of DSPN. For the two most frequently abnormal attributes (fibular MNCV and sural SNAP), the agreement was not as good (kappa coefficient = 0.29) as it was for fibular MNCV and tibial MNCV (kappa = 0.57). Because high kappa scores indicate that the two attributes are measuring the same dysfunction, their use in a composite score may not increase sensitivity as much as other pairs (e.g., fibular MNCV and sural SNAP) (Table 1).

**Performance of Different Criteria for DSPN.** Criterion 1 was the presence of ≥1 abnormality of any one attribute of NC from among all 14 attributes evaluated (the criterion used by some electromyographers). This criterion uses an inexact prespecification of abnormalities; any one of many attributes of several nerves could be abnormal. The abnormalities found may or may not be representative of DSPN. Use of this criterion results in a large type 1 error (i.e., overestimating disease due to multiple measurements; that is, 37.3%, 17.3%, and 9.1% from use of ≤5th/≥95th, ≤2.5th/≥97.5th, or ≤1st/≥99th percentile criteria) (Table 1).

Criterion 2 involved the presence of “≥1 abnormal attributes in ≥2 separate nerves tested” (an early RDNS criterion6,17–19). Like criterion 1, different NC attribute abnormalities could be identified. In healthy subjects, the frequency of abnormality is overestimated using the ≥5th/≥95th criterion (11.2%) (Table 1). Sensitivity in the RDNS cohort is high (i.e., 34.6% and 22.1% for ≤2.5th/≥97.5th and ≤1st/≥99th percentiles). Criterion 3 was “an abnormality (≤1st or ≥99th percentile of any attributes of NG of two separate nerves, one of which must be the sural nerve)” — abnormality of signs and symptoms were not considered in this analysis (AAN, American
Association of Neuromuscular and Electrodiagnostic Medicine, and American Academy of Physical Medicine and Rehabilitation consensus criteria). One of the two criteria is precisely defined. In the HS cohort, there was an overestimation of affected persons using <5th/≥95th criteria (7.3%) in the RDNS-HS. In the DM cohort, sensitivity for ≤1st/≥99th was low (15.6%) when compared with composite scores 4–8.

Criterion 4 was abnormality of each of the two most sensitive attributes of NC observed in the RDNS cohort (i.e., fibular MNCV and sural SNAP) and at defined percentile abnormalities. Use of this criterion results in low specificity values in the HS and low sensitivity in the DM cohort.

Criteria 5–8 are described together because of their similar performance in both the HS and the DM cohort. Criterion 5 combines the two most sensitive attributes of NC into a composite score (i.e., fibular MNCV and sural SNAP). Criterion 6 is a composite score combining the two attributes with the highest agreement for the diagnosis of DSPN (i.e., fibular MNCV and tibial MNCV). Criterion 7 is abnormality of a composite score of attributes of NC considered to be representative of DSPN (i.e., Σ 5 NC nds consists of fibular CMAP, MNCV and MNDL, tibial MNDL, and sural SNAP). Criterion 8 is a composite score of the six most sensitive attributes of NC abnormality observed in the RDNS cohort (i.e., fibular MNCV, sural SNAP, tibial MNCV, ulnar MNCV, tibial F-latency, and ulnar F-latency).

All these composite scores (5–8) provided close estimates of the set specificities in the HS cohort and high sensitivity in the DM cohort (Table 1). The highest sensitivity was achieved by criterion 5, with values of 37.9% and 31.1% for ≤2.5th/97.5th and ≤1st/99th percentiles.

**DISCUSSION**

In this study we have assessed the databases of previously studied cross-sectional cohorts of population-representative healthy subjects (RDNS-HS) and patients with diabetes mellitus (RDNS) and addressed three questions: (1) At a given percentile level of abnormality (i.e., ≤2.5th/97.5th), which are the most sensitive attributes of NC abnormality and useful for the diagnosis of DSPN? (2) What is the agreement between pairs of attributes of NC for the diagnosis of DSPN? (3) Among eight NC criteria for the diagnosis of DSPN, which perform best by the criteria of being representative of NC abnormality characteristic of DSPN, avoidance of type 1 error from assessment of multiple measurements, obtaining high degrees of specificity in a healthy subject cohort, and high degrees of sensitivity in a diabetic cohort? Our study focused on patients with typical DPN (i.e., DSPN), because atypical DPN (intercurrent small-fiber sensory and autonomic neuropathies) may not be detectable using NC approaches.

We found that the six most sensitive attributes of NC in the RDNS cohort in decreasing frequencies were fibular MNCV, sural SNAP, tibial MNCV, ulnar MNCV, tibial F-wave latency, and ulnar F-wave latency, using, as the basis for comparison, abnormality criteria at the ≤2.5th/97.5th percentiles, as estimated from study of the RDNS-HS cohort. We have previously studied the clinometric performance of attributes of NC versus that of other neurophysiological and clinical tests in the diagnosis of DSPN,14,15,17,19,21,22

This study of different criteria for DSPN provides several lines of useful information and finds that it provides objective and valid diagnostic information. First, it shows that not all attributes are equally sensitive for the diagnosis of DSPN. Second, rates of abnormality from use of any single attribute are below that of use of composite NC endpoints. These inferences, however, apply only to DSPN, not to atypical DPN or to focal or multifocal varieties.

The degree of agreement between pairs of NC tests, using the kappa coefficient, provides additional useful information on which attribute might be useful as components of composite NC scores for the diagnosis of DSPN. Although it was considered to be desirable that pairs of attributes should show significant agreement (thus indicative that both were an indication of DSPN), we found that high degrees of agreement did not necessarily improve their performance in composite scores. This is illustrated by comparison of fibular MNCV and tibial MNCV vs. fibular MNCV and sural SNAP—the first pair has a higher degree of agreement, but when combined as a composite score lower sensitivity (Tables 1 and 2).

The eight NC criteria for the diagnosis of DSPN provided quite different estimates of abnormality in both our HS and DM cohorts. A critical examination of the most valid criteria is therefore justified. Criterion 1 is an unacceptable criterion for several reasons. The abnormality could be due to a mononeuropathy and not a polyneuropathy. Second, the abnormalities are not specifically defined. Many different outcomes could result from its use; therefore, it is unsuitable for epidemiological surveys or randomized clinical trials. An elevated false positive rate was to be expected and was demonstrated in our RDNS-HS cohort due to the large number of attributes studied (a corollary to the elevated type 1 error rate that occurs when multiple-hypothesis testing is performed). In particular, in a healthy subject cohort with values
normally distributed and the attributes not linked to each other (i.e., statistically independent), 46%, 26%, and 11% of patients would be classified as DSPN when only 5%, 2.5%, and 1% would be expected to show abnormality. Empirical evaluation of the RDNS-HS cohort resulted in frequencies of 37%, 17%, and 9%—somewhat lower values than the theoretical and suggestive of correlation, or clustering, of some of the 14 attributes (Table 1). These high false positive determinations of DSPN would be detrimental to diagnostic utility not only in research but also in medical practice. Because of the lack of specificity using this criterion, excessively high estimates of DSPN were found in RDNS (Table 1).

Criterion 2 performed considerably better than criterion 1. Use of this criterion ensures that a mononeuropathy is highly unlikely to explain the abnormality. Like criterion 1, specificity is higher than it should be in RDNS-HS using the 5th percentile abnormality criterion. This criterion is more stringent than criterion 1, so it is expected to provide greater control of the false positive rate. Using \( \leq 2.5\text{th}/\geq 97.5\text{th} \) or \( \leq 1\text{st}/\geq 99\text{th} \) percentiles as the threshold for abnormality, very low rates of abnormality were observed in the HS cohort; however, the criteria could produce estimates of the prevalence of DSPN that are highly sensitive to the threshold selected. In particular, shifting from the \( \leq 2.5\text{th}/\geq 97.5\text{th} \) to the more stringent \( \leq 1\text{st}/\geq 99\text{th} \) percentiles decreases the estimated frequency of DSPN from 34.7% to 22.1% (Table 1).

Criterion 3 did not perform well if followed. As compared with criterion 2, there was prespecification of one attribute (sural SNAP), and percentile abnormality is set at the \( \leq 1\text{st} \) percentile, but there was failure to exactly prespecify other attributes. In the RDNS-HS cohort, abnormality using this criterion occurred in <1% of patients (0.6%) and resulted in a low (too low considering frequencies reported in the medical literature and composite score estimates) prevalence of DSPN (e.g., using the AAN criteria, it was 15.6% in the RDNS cohort). If the \( \leq 2.5\text{th} \) or \( \geq 97.5\text{th} \) percentile thresholds were used, the estimated frequency of DSPN in RDNS would increase to 26.3%—a more reasonable estimate. As for criterion 2, the percentile cut-off used (i.e., \( \leq 2.5\text{th}/\geq 97.5\text{th} \) or \( \leq 1\text{st}/\geq 99\text{th} \)) had a major effect on frequency of DSPN in the RDNS cohort.

Criterion 4 has exact pre-specified criteria and of percentile abnormality, but it performed poorly. In RDNS-HS it resulted in an insufficiently low frequency of abnormality (false positive rates of 0.6% and 0.3%) for \( \leq 2.5\text{th} \) and \( \leq 1\text{st} \) percentiles. For the DSPN cohort, the criterion resulted in too low an estimate of DSPN (i.e., sensitivity of 11.8% and 3.9%)—well below expected frequencies of DSPN in a population-based cohort of patients with DM and also well below the figures obtained using composite scores 5–8.

The four composite scores of NC (criteria 5–8) performed very well in both the RDNS-HS and RDNS cohorts. All four composite scores were highly specific in the HS cohort with a frequency of abnormality near the specificity criteria of \( \leq 5\text{th}/\geq 95\text{th} \), \( \leq 2.5\text{th}/\geq 97.5\text{th} \), or \( \leq 1\text{st}/\geq 99\text{th} \) percentiles. In the RDNS cohort, however, sensitivity in the strict diagnostic accuracy context could not be assessed. Assuming that specificity is the same for the RDNS, as it was shown to be in the RDNS-HS cohort (not an unreasonable assumption, because the technique of testing and the criteria for abnormality are exactly alike), sensitivity can be assessed in the RDNS. Therefore, our results support the idea that these four composite scores are both specific and sensitive for the accurate detection of DSPN. As anticipated, their use did not overdiagnose DSPN in the HS cohort; they provided abnormalities very near the prespecified percentile levels. It is also important to note that the attributes chosen for composite scores 5–8 are all thought to be representative of nerve dysfunction in DSPN (i.e., leg nerves) and inclusive of different functional neurophysiological abnormalities. The use of composite scores allows inclusion of disparate functional attributes, avoids identification of false positives, and, because they are continuous measurements, allows estimates of function and dysfunction through the range of normality and abnormality. Therefore, composite scores may be used to assess latent neuropathic dysfunction and severity of DSPN, a clear advantage over dichotomous judgments. The good quality of composite scores makes them especially useful for epidemiological surveys and randomized, controlled trials, a subject we have discussed previously. Considering their good qualities, why are composite scores not used more widely? It is possible that the needed reference values, as corrected for by the applicable variables, are not generally available in a readily usable form, and some degree of mathematical manipulation of the data is needed. Such reference databases and computational facilities could be made available for some ethnic and geographic cohorts and probably should be used more widely. Software programs could be made available by a medical institution, a governmental agency, or by an equipment manufacturer.

Of the four criteria assessed, criteria 1 and 4 performed unacceptably. Despite some shortcomings (i.e., some degree of false positives and lack of complete prespecification), criteria 2 and 3
performed reasonably well, but only when the \( \leq 2.5\text{th}/\geq 97.5\text{th} \) percentile pairing was used. Overall, the composite scores performed best.

The authors thank Jenny Davies, BA, for statistical analysis and Mary Lou Hunziker for preparation of the manuscript. This study was supported in part by grants obtained from the National Institutes of Neurological Disorders and Stroke (NS36797) and the Mayo Clinic Center for Clinical and Translational Research (U54RR 24150-1).

REFERENCES


