Published in final edited form as:

Arch Neurol. 2012 December 1; 69(12): 1609–1614. doi:10.1001/archneurol.2012.1481.

“Unequivocally Abnormal” vs “Usual” Signs and Symptoms for Proficient Diagnosis of Diabetic Polyneuropathy:

CI vs N Phys Trial

Dr. Peter J. Dyck, MD, Ms. Carol J. Overland, Dr. Phillip A. Low, MD, Dr. William J. Litchy, MD, Ms. Jenny L. Davies, BA, Dr. P. James B. Dyck, MD, Dr. Rickey E. Carter, PhD, Dr. L. Joseph Melton III, MD, Dr. Henning Andersen, MD, Dr. James W. Albers, MD, Dr. Charles F. Bolton, MD, Dr. John D. England, MD, Dr. Christopher J. Klein, MD, Dr. Gareth Llewelyn, MD, Dr. Michelle L. Mauermann, MD, Dr. James W. Russell, MD, Dr. Dinesh Selvarajah, MD, Dr. Wolfgang Singer, MD, Dr. A. Gordon Smith, MD, Dr. Solomon Tesfaye, MD, and Dr. Adrian Vella, MD

Peripheral Neuropathy Research Laboratory, Department of Neurology (Drs P. J. Dyck, Low, Litchy, P. J. B. Dyck, Klein, Mauermann, and Singer and Mss Overland and Davies), and Divisions of Biostatistics (Dr Carter), Epidemiology (Dr Melton), and Endocrinology (Dr Vella), Mayo Clinic, Rochester, Minnesota; Departments of Neurology, Aarhus University Hospital, Aarhus, Denmark (Dr Andersen), University of Michigan, Ann Arbor (Dr Albers), Queen's University, Kingston, Ontario, Canada (Dr Bolton), Louisiana State University, New Orleans (Dr England), University Hospital of Wales, Cardiff, England (Dr Llewelyn), University of Maryland, Baltimore (Dr Russell), and University of Utah, Salt Lake City (Dr Smith); and Departments of Human Metabolism (Dr Selvarajah) and Diabetes (Dr Tesfaye), Sheffield Teaching Hospitals, Sheffield, England

Abstract

Objective—To repeat the Clinical vs Neurophysiology (CI vs N Phys) trial using “unequivocally abnormal” signs and symptoms (Trial 2) compared with the earlier trial (Trial 1), which used “usual” signs and symptoms.

Design—Standard and referenced nerve conduction abnormalities were used in both Trials 1 and 2 as the standard criterion indicative of diabetic sensorimotor poly-neuropathy. Physician proficiency (accuracy among evaluators) was compared between Trials 1 and 2.

Setting—Academic medical centers in Canada, Denmark, England, and the United States.

© American Medical Association. All rights reserved

Correspondence: Peter J. Dyck, MD, Peripheral Neuropathy Research Laboratory, Department of Neurology, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (dyck.peter@mayo.edu).

Conflict of Interest Disclosures: None reported.

Additional Contributions: We received help from personnel in the Peripheral Neuropathy Research Laboratory, Electromyographic Laboratory, Quantitative Sensation Testing Laboratory, Autonomic Laboratory, and Section of Engineering at Mayo Clinic. Mary Lou Hunziker assisted with preparation of the manuscript.

Author Contributions: Drs P. J. Dyck and Ms Davies had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* P. J. Dyck, Low, Litchy, P. J. B. Dyck, Bolton, England, Mauermann, Russell, and Tesfaye. *Acquisition of data:* Overland, Litchy, Andersen, Albers, England, Klein, Llewelyn, Mauermann, Selvarajah, Singer, Smith, and Vella. *Analysis and interpretation of data:* Low, Davies, Carter, Melton, Andersen, Albers, England, Selvarajah, Singer, Smith, Tesfaye, and Vella. *Drafting of the manuscript:* Low, Carter, Russell, and Selvarajah. *Critical revision of the manuscript for important intellectual content:* P. J. Dyck, Overland, Low, Litchy, Davies, P. J. B. Dyck, Carter, Melton, Albers, Bolton, England, Klein, Llewelyn, Mauermann, Russell, Selvarajah, Singer, Smith, Tesfaye, and Vella. *Statistical analysis:* Davies, Carter, and Selvarajah. *Administrative, technical, and material support:* Overland, Litchy, Albers, Klein, Russell, Singer, Smith, and Vella. *Study supervision:* Low, Andersen, and Vella.

Participants—Thirteen expert neuromuscular physicians. One expert was replaced in Trial 2.

Results—The marked overreporting, especially of signs, in Trial 1 was avoided in Trial 2. Reproducibility of diagnosis between days 1 and 2 was significantly ($P=.005$) better in Trial 2. The correlation of the following clinical scores with composite nerve conduction measures spanning the range of normality and abnormality was improved in Trial 2: pinprick sensation ($P=.03$), decreased reflexes ($P=.06$), touch-pressure sensation ($P=.06$), and the sum of symptoms ($P=.06$).

Conclusions—The simple pretrial decision to use unequivocally abnormal signs and symptoms—taking age, sex, and physical variables into account—in making clinical judgments for the diagnosis of diabetic sensorimotor polyneuropathy (Trial 2) improves physician proficiency compared with use of usual elicitation of signs and symptoms (Trial 1); both compare to confirmed nerve conduction abnormality.

Neuropathy signs and symptoms are used to diagnose and scale the severity of distal symmetric sensorimotor polyneuropathy, such as the typical variety associated with diabetes mellitus, diabetic sensorimotor polyneuropathy (DSPN). One usually assumes that physicians can accurately elicit signs and symptoms and, with their use, diagnose DSPN sensitively and accurately and judge its severity.^{1–5} This assumption had not been rigorously tested until recently,^{6,7} perhaps because the use of signs and symptoms for this purpose had been endorsed by expert consensus panels.^{5,8,9} In the first of 2 studies addressing the accuracy of signs and symptoms for this purpose, 3 neurologists from the Mayo Clinic independently examined 20 patients with diabetes mellitus without and with DSPN on 2 occasions and judged the scored abnormality of neuropathic signs and symptoms.^{6,10} These researchers found a high degree of reproducibility of their clinical judgments, and clinical assessments were closely correlated with nerve conduction (NC) and abnormal quantitative sensation test results. However, quite different results were obtained in the CI vs N Phys Trial 1 (Trial 1), the forerunner of the present CI vs N Phys Trial 2 (Trial 2), in which 12 expert international physicians examined each of 24 masked patients with diabetes mellitus without and with DSPN on 2 consecutive days, without conferring among them-selves or receiving instruction as to how they should elicit or judge neuropathy signs or symptoms or diagnose DSPN.⁷ In Trial 1, the physicians used their “usual” presumably sensitive criteria for eliciting signs, symptoms, and clinical diagnosis of DSPN.⁷ Individual physician test-retest reproducibility (κ value) of signs, symptoms, and diagnosis were generally good to very good, but judgments were excessively variable and inaccurate among physicians when compared with a highly standardized and referenced composite standard normal deviate score of 5 attributes of NC abnormality ($\Sigma 5$ NC nds 2.5th percentile). Poor physician proficiency (accuracy among investigators) was attributed to significant excessive and incorrect judging of clinical signs and, to a lesser degree, of symptoms and diagnosis. Recognizing that even experts are not as proficient as desired, we decided to retest proficiency using more specific pretrial criteria for eliciting signs and symptoms. Whereas in Trial 1 expert physicians had used usual presumably sensitive elicitation of signs and symptoms, in Trial 2 they agreed to use only “unequivocal” or certain abnormal signs and symptoms (presumably more specific than usual criteria) indicative of DSPN, with age, sex, physical fitness, and physical variables taken into account in judging abnormality.

To assess physician proficiency, an independent criterion standard indicator was needed, not only for the occurrence but also for the severity of DSPN. As reviewed in the “Comment” section, a composite score of NC ($\Sigma 5$ NC nds) was used for this purpose.¹¹

Therefore, the primary question addressed in Trial 2 is whether physician proficiency is improved by more specifically eliciting unequivocally abnormal signs and symptoms for the

clinical diagnosis of DSPN compared with the usual elicitation of signs and symptoms in Trial 1. We compared both results with standard composite NC abnormality.

METHODS

DESIGN AND CONDUCT OF TRIAL 2

In Trials 1 and 2, using only elicited signs and symptoms, 12 expert neuromuscular physicians examined 24 masked patients (disguised by wearing surgical garments, caps, and masks and by using electronic distortion of their voices [Figure 1]) and judged and recorded the signs and symptoms of DSPN. In Trial 1, usual (presumably sensitive) criteria were used for eliciting signs and symptoms.⁷ In Trial 2 (the present report), unequivocally abnormal signs and symptoms with correction for applicable variables of age, sex, physical fitness, and physical features were used.

In Trial 1, physicians had not been instructed on how to examine, grade, or judge the signs or symptoms or how to diagnose DSPN, and they were asked not to confer among themselves as to how they evaluated or judged abnormal findings or diagnosed DSPN. They were to use their usual approaches and criteria for the clinical assessment of signs and symptoms and for the clinical diagnosis of DSPN. After clinical examination of each patient in both trials, physicians recorded the presence or absence of signs (muscle weakness, hyporeflexia or areflexia, and decreased sensation [touch pressure, vibration, joint motion, pinprick, or other]) and symptoms (weakness, positive neuropathic sensory symptoms of “asleep-numbness,” “prickling,” or pain [“deep aching,” “lancinating,” “burning,” or “other”]). Forms were provided for this purpose. For both trials, NC assessments were performed within 1 or 2 weeks of the trial to assess fibular nerve motor amplitude, velocity and distal latency, tibial motor distal latency, and sural sensory nerve amplitude. Abnormality of NC was assessed using Σ 5 NC nds at the 2.5th percentile or less as previously described.¹¹

The design and execution of Trial 2 is exactly like that of Trial 1 with the following 3 differences: (1) Trial 2 was performed 1 year after Trial 1; (2) 1 expert physician and 3 patients from Trial 1 were not able to participate in Trial 2, necessitating substitutions; and (3) a consensus process preceded conduct of Trial 2, in which physicians agreed to grade as abnormal only unequivocal (certain) abnormality of signs, symptoms, and diagnosis, taking age, sex, physical fitness, height, and weight into consideration before judging signs and symptoms as abnormal. The masking of patients' identity and medical conditions was otherwise alike between Trials 1 and 2. Masking of each patient's identity and disease condition and the individual physician's performance were maintained within and between the trials and to the present.

As previously reported, the 12 expert neuromuscular physicians came from 4 countries (Canada, Denmark, England, and the United States). The 24 volunteers with diabetes mellitus without and with DSPN for Trial 1 had been recruited using NC criteria so that approximately half of them did not have DSPN and half did to varying degrees of severity. None of the patients had selective small-fiber painful polyneuropathy (ie, chronic idiopathic axonal polyneuropathy or atypical diabetic polyneuropathy). The NC criterion standard for DSPN was Σ 5 NC nds at the 2.5th percentile or less and was the same for Trials 1 and 2. The criterion standard of at least 75% of physician diagnosis of DSPN was not used in this study because of its poor performance in Trial 1.⁷

STATISTICAL ANALYSIS

The primary question addressed in this comparison of Trials 1 and 2 was the proficiency of 12 expert neuromuscular physicians to diagnose DSPN when they used usual (Trial 1) vs

unequivocally abnormal criteria (Trial 2; more specific criteria) to elicit signs and symptoms for the diagnosis of DSPN using abnormality of NC as the objective and standard indication of DSPN. We used standard statistical tests to compare frequency distributions, reproducibility, and correlation of frequency of signs and symptoms compared with NC abnormality in Trial 2 vs Trial 1.

RESULTS

TEST-RETEST REPRODUCIBILITY OF SIGNS, SYMPTOMS, AND CLINICAL DIAGNOSIS OF DSPN IN TRIAL 2 VS TRIAL 1

In both trials, reproducibility of each physician's assessment of signs, symptoms, and diagnoses between days 1 and 2 ranged from good to very good and, for most physicians, was statistically significant (Table 1). Using the Wilcoxon signed rank test, a large and significant improvement in reproducibility of diagnosis (κ value) was found in Trial 2 compared with Trial 1 (median [range], 0.80 [0.55–0.88] in Trial 2 vs 0.49 [0.19–0.75] in Trial 1 [$P=.005$]). Reproducibility of judgment of the signs or symptoms was not significantly increased in Trial 2.

TEST-RETEST REPRODUCIBILITY OF NERVE CONDUCTION ABNORMALITY AND REPRODUCIBILITY OF COMPOSITE NERVE CONDUCTION SCORE

Test-retest reproducibility of NC abnormality (Σ 5 NC nds $\geq$ 2.5th percentile) for the diagnosis of DSPN (a dichotomous measure) between days 1 and 2 was 0.83 in Trial 1 and 0.88 in Trial 2, both highly reproducible. Reproducibility of the composite score of NC (Σ 5 NC nds) as a continuous measure of severity spanning normality and abnormality was assessed using intraclass correlation coefficients. The intraclass correlation coefficient was 0.93 for Trial 1 and 0.96 for Trial 2, indicating high reproducibility of this measure in both trials.

OVERREPORTING OF SIGNS, SYMPTOMS, AND CLINICAL DIAGNOSIS OF DSPN

On the assumption that confirmed NC abnormality (Σ 5 NC nds $\geq$ 2.5th percentile) is a sensitive, objective, and quantitative indication of DSPN (see the "Comment" section), the degree in the overreporting of signs, symptoms, and clinical diagnosis was used to assess for accuracy of the clinical evaluation. The remarkable improvement of overreporting of signs, symptoms, and diagnosis in Trial 2 compared with Trial 1 is shown in Figure 2. Trial 1 had a large degree of overreporting, especially of signs (45% and 47% of 576 physician evaluations on days 1 and 2, respectively), with less overreporting of symptoms (14% and 12%, respectively) and clinical diagnosis (26% and 27%, respectively) (Figure 2). Also in Trial 1, overreporting of at least 75% of physicians' judgments of abnormal signs (9 of 24 and 11 of 24 on days 1 and 2, respectively) occurred. However, overreporting of at least 75% of physicians' assessments of symptoms was found in lesser frequencies (ie, 2 of 24 and 3 of 24 and for 75% clinical diagnosis in 1 of 24 and 3 of 24, respectively) (Table 2). By contrast, in Trial 2, individual physician overreporting of signs occurred in only 5% and 5%, with lesser percentages of overreporting of symptoms and diagnosis (Figure 2). No overreporting of at least 75% of physicians' judgments of signs, symptoms, or diagnosis occurred in Trial 2 (Table 2). The markedly reduced overreporting of at least 75% of physicians' judgments of signs in Trial 2 was highly significant (ie, $P=.005$ on day 1 and $P=.003$ on day 2) (Table 2).

CORRELATION OF PHYSICIANS' SCORING OF CLINICAL SIGNS AND SYMPTOMS AND NERVE CONDUCTION

On days 1 and 2 of Trial 1, 9 of 12 and 6 of 12 physicians, respectively, had a statistically significant correlation between their crude scores of elicited signs and symptoms and continuous NC measures (Σ 5 NC nds) (Table 3). In Trial 2, 12 of 12 examiners had significant correlations on days 1 and 2. For day 2, this difference between the trials was almost significant ($P = .06$).

Physicians' assessments of individual signs and symptoms compared with the continuous measure of nerve function were compared between Trials 2 and 1 using similar approaches to those listed in Table 3. Assessment of individual symptoms was not significantly improved in Trial 2 compared with Trial 1; however, the sum of elicited symptoms almost reached significance ($P = .06$). Of the elicited signs, pinprick sensation ($P = .03$), touch-pressure sensation ($P = .06$), and decreased reflexes ($P = .06$) resulted in better performance in Trial 2.

COMMENT

Evaluation of signs and symptoms is widely used for medical diagnosis, for example, of disease conditions such as typical DSPN. Considering their widespread use, surprisingly little reliable data are available on how proficiently these evaluations are performed. This fact was emphasized in a recent editorial by Johnston and Hauser.¹² Reflecting on the "beautiful and ethereal" neurological examination, they state that "little evidence actually supports its value." They note further that the examination "is easy to do poorly, and we know very little about how well it performs in the hands of experts, much less many other practitioners."¹²

In the previously reported Trial 1,⁷ expert physicians relying on their elicited signs and symptoms alone were not proficient in diagnosing DSPN. Using their usual presumably sensitive elicitation of signs, symptoms, and diagnosis, excessive variability and inaccuracy were found compared with confirmed NC abnormality. Although results of their clinical assessments were reproducible, the experts markedly overreported signs in particular compared with NC abnormality.

In Trial 2 (essentially a repeat of Trial 1 one year later), almost the same physicians examined almost the same masked patients using unequivocally abnormal signs and symptoms and took variables of age, sex, physical fitness, and physical characteristics into account. Using the more specific Trial 2 criteria, physician proficiency was significantly less variable and more accurate in Trial 2 than in Trial 1 when both evaluations were compared with confirmed NC abnormality.

The choice of Σ 5 NC nds at the 2.5th percentile or less as the dichotomous and Σ 5 NC nds as the continuous criterion standard measure of DSPN probably needs further explanation and defense. We used these measures because they are among the most sensitive, objective, and quantitative independent measures against which clinical signs, symptoms, and diagnosis could be compared. This conclusion is based on results of cohort studies,^{1,13-15} epidemiologic surveys,^{6,7,11,15-17} and consensus panel reports.^{5,8,9,17} This conclusion is further confirmed by the present results. The composite NC score chosen for these studies combines the attributes of 3 leg nerves and 4 functions (known to be affected in DSPN), and the percentile abnormality was set by comparison with a reference cohort of 330 healthy subjects from Olmsted County, Minnesota. As shown in a previous study from this group,¹¹ use of this composite score also avoids type I error. In the present study, the chosen composite NC score was shown to be a highly reproducible indication of DSPN. Therefore,

this composite NC score can be used for the sensitive diagnosis of DSPN but also as a continuous measure of nerve function spanning normality and abnormality.

The improved accuracy of assessment of signs, symptoms, and diagnosis in Trial 2 compared with Trial 1 likely relates to more specific elicitation of signs and symptoms in Trial 2. We make this inference because more specifically elicited signs and symptoms were the main differences between the design and conduct of Trials 2 and 1. Also, physician recall of patients' history or findings from Trial 1, a possible explanation for improved proficiency in Trial 2, should not have occurred because each patient's identity, clinical findings, and diagnosis remained masked. Furthermore, using more obvious "harder" signs or symptoms as criteria would not necessarily improve proficiency because, although obvious cases might more readily be identified, milder, less obvious cases might go unidentified. Therefore, we can infer that the improved proficiency found in Trial 2 results from the use of more specific evaluation of signs and symptoms, taking such factors as age, sex, and physical variables into account when judging the signs, symptoms, and diagnosis.¹⁸

By using unequivocally abnormal signs and symptoms as in Trial 2, have the signs, symptoms, and diagnosis of DSPN been underestimated? Underestimation is possible but seems unlikely to be of a considerable magnitude because, even with the use of the more specific criteria for DSPN, a small percentage of overreporting of signs, symptoms, and clinical diagnosis occurred.

An additional possible insight coming from these trials relates to the greater overreporting of signs compared with symptoms. Because symptoms are reported by patients and signs by physicians, it could be inferred that patients' judgments are more accurate than those of physicians. However, when physicians used the more specific Trial 2 criteria, a reasonable balance among signs, symptoms, and diagnosis was achieved. Assuming that insights arrived at in this report can be generalized, we might reasonably conclude that specific rather than sensitive elicitation of signs and symptoms should be advocated to improve accurate and proficient elicitation of signs and symptoms for the diagnosis of DSPN. This insight might be applied to conduct of therapeutic trials, neurological education, and medical practice.

Acknowledgments

Funding/Support: This study was supported in part by the Kawamura Fund, by the Society for the Support of Neurologic Education and Research of Rochester, by grant NS36797 from the National Institutes of Neurological Disorders and Stroke, by grant AG034676-47 from the Rochester Epidemiology Project (principle investigator, Walter A. Rocca, MD), and by Mayo Foundation funds.

References

1. Gilliatt, RW. Clinical aspects of diabetic neuropathy. In: Cumings, JN.; Kremer, M., editors. *Biochemical Aspects of Neurological Disorders*. Oxford, England: Blackwell; 1965. p. 117-142.
2. Pirart J. Diabetic neuropathy: a metabolic or a vascular disease? *Diabetes*. 1965; 14:1-9. [PubMed: 14260096]
3. Nilsson SV, Nilsson JE, Frostberg N, Emilsson T. The Kristianstad survey, II: studies in a representative adult diabetic population with special reference to comparison with an adequate control group. *Acta Med Scand Suppl*. 1967; 469:1-42. [PubMed: 5231220]
4. Franklin GM, Kahn LB, Baxter J, Marshall JA, Hamman RF. Sensory neuropathy in non-insulin-dependent diabetes mellitus: the San Luis Valley Diabetes Study. *Am J Epidemiol*. 1990; 131(4): 633-643. [PubMed: 2316495]
5. England JD, Gronseth GS, Franklin G, et al. American Academy of Neurology; American Association of Electrodiagnostic Medicine; American Academy of Physical Medicine and

Rehabilitation. Distal symmetric polyneuropathy: a definition for clinical research: report of the American Academy of Neurology, the American Association of Electrodiagnostic Medicine, and the American Academy of Physical Medicine and Rehabilitation. *Neurology*. 2005; 64(2):199–207. [PubMed: 15668414]

6. Dyck PJ, Kratz KM, Lehman KA, et al. The Rochester 54 Neuropathy Study: design, criteria for types of neuropathy, selection bias, and reproducibility of neuropathic tests. *Neurology*. 1991; 41(6):799–807. [PubMed: 2046920]
7. Dyck PJ, Overland CJ, Low PA, et al. CI vs NPhys Trial Investigators. Signs and symptoms versus nerve conduction studies to diagnose diabetic sensorimotor polyneuropathy: CI vs NPhys trial. *Muscle Nerve*. 2010; 42(2):157–164. [PubMed: 20658599]
8. Report and recommendations of the San Antonio conference on diabetic neuropathy. *Neurology*. 1988; 38(7):1161–1165. [PubMed: 3386839]
9. Tesfaye S, Boulton AJ, Dyck PJ, et al. Toronto Diabetic Neuropathy Expert Group. Diabetic neuropathies: update on definitions, diagnostic criteria, estimation of severity, and treatments. *Diabetes Care*. 2010; 33(10):2285–2293. [PubMed: 20876709]
10. Dyck PJ, Sherman WR, Hallcher LM, et al. Human diabetic endoneurial sorbitol, fructose, and myo-inositol related to sural nerve morphometry. *Ann Neurol*. 1980; 8(6):590–596. [PubMed: 7212646]
11. Dyck PJ, Carter RE, Litchy WJ. Modeling nerve conduction criteria for diagnosis of diabetic polyneuropathy. *Muscle Nerve*. 2011; 44(3):340–345. [PubMed: 21996793]
12. Johnston SC, Hauser SL. The beautiful and ethereal neurological exam: an appeal for research. *Ann Neurol*. 2011; 70(2):A9–A10. [PubMed: 21823150]
13. Mulder DW, Lambert EH, Bastron JA, Sprague RG. The neuropathies associated with diabetes mellitus: a clinical and electromyographic study of 103 unselected diabetic patients. *Neurology*. 1961; 11(4 pt 1):275–284. [PubMed: 13773672]
14. Albers JW, Herman WH, Pop-Busui R, et al. Diabetes Control and Complications Trial/ Epidemiology of Diabetes Interventions and Complications Research Group. Effect of prior intensive insulin treatment during the Diabetes Control and Complications Trial (DCCT) on peripheral neuropathy in type 1 diabetes during the Epidemiology of Diabetes Interventions and Complications (EDIC) Study. *Diabetes Care*. 2010; 33(5):1090–1096. [PubMed: 20150297]
15. Dyck PJ, O'Brien PC. Polyneuropathy dysfunction scores. *J Neurol Neurosurg Psychiatry*. 2006; 77(8):899–900. [PubMed: 16709581]
16. Dyck PJ, Kratz KM, Karnes JL, et al. The prevalence by staged severity of various types of diabetic neuropathy, retinopathy, and nephropathy in a population-based cohort: the Rochester Diabetic Neuropathy Study. *Neurology*. 1993; 43(4):817–824. [PubMed: 8469345]
17. Dyck PJ, Bushek W, Spring EM, et al. Vibratory and cooling detection thresholds compared with other tests in diagnosing and staging diabetic neuropathy. *Diabetes Care*. 1987; 10(4):432–440. [PubMed: 3622200]
18. Dyck, PJ.; Hughes, RAC.; O'Brien, PC. Quantitating overall neuropathic symptoms, impairments, and outcomes. In: Dyck, PJ.; Thomas, PK., editors. *Peripheral Neuropathy*. 4. Philadelphia, PA: Elsevier; 2005. p. 1031-1052.

Figure 1. The clothing, masks, and dark glasses used to disguise the identity of patients (only 12 of the 24 patients are seen in the front row). The voices of patients were electronically altered. Patients did not speak unless physicians wore earphones. Some study personnel are also shown.

Figure 2.

The frequency of at least 75% of expert physicians' elicitation of signs, symptoms, and clinical diagnosis of diabetic sensorimotor polyneuropathy. Confirmed nerve conduction (NC) abnormality (calculated as the composite standard normal deviate score of 5 attributes of NC abnormality [$\Sigma 5$ NC nds ≤ 2.5 th percentile]) is compared with results on days 1 and 2 of the CI vs N Phys Trial 1 (A) and Trial 2 (a replication of Trial 1) (B). Percentages are calculated from 576 physician evaluations for 24 patients. Ellipses indicate not applicable.

*Overreported clinical diagnoses were defined as the presence of clinical abnormalities but no NC abnormalities.

Test-Retest Reproducibility of Expert Neuromuscular Physicians' Judgments of Signs, Symptoms, and Diagnosis of DSPN Between Days 1 and 2 of CI vs N Phys Trials 1 and 2^a

Table 1

Physician No.	Trial 1				Trial 2			
	Signs		Symptoms		Signs		Symptoms	
	κ	P Value	κ	P Value	κ	P Value	κ	P Value
1	0.41	.02	1.00	<.001	0.50	.006	0.57	.003
2 ^b	0.75	<.001	0.83	<.001	0.47	.01		
3	0.65	<.001	0.55	.003	0.49	.009	0.71	<.001
4	0.51	.002	0.73	<.001	0.57	.001	0.74	<.001
5	1.00	<.001	0.68	<.001	0.49	.008	0.66	<.001
6	0.42	.01	0.66	<.001	0.49	.008	0.65	<.001
7	0.59	.002	0.71	<.001	0.33	.04	0.78	<.001
8	0.62	.001	0.22	.12	0.30	.04	0.71	<.001
9	0.48	.006	0.81	<.001	0.75	<.001	0.63	<.001
10	0.78	<.001	0.67	<.001	0.63	<.001	0.34	.04
11	1.00	<.001	0.75	<.001	0.19	.17	0.49	.009
12	0.36	.04	0.88	<.001	0.66	<.001	0.58	.002
13 ^b							0.64	<.001

Abbreviation: DSPN, diabetic sensorimotor polyneuropathy.

^a Boldface type indicates statistically significant correlations. Wilcoxon signed rank tests comparing Trials 1 and 2 were not significantly different for signs (median [range], 0.60 [0.36–1.00] for Trial 1 and 0.65 [0.34–0.78] for Trial 2 [$P = .83$]) and for symptoms (0.72 [0.22–1.00] for Trial 1 and 0.65 [0.28–1.00] for Trial 2 [$P = .32$]). Differences were significant for diagnosis (0.49 [0.19–0.75] for Trial 1 and 0.80 [0.55–0.88] for Trial 2 [$P = .005$]).

^b Physician 2 was unavailable for Trial 2 and was replaced by physician 13.

Correlations of Composite Scores of Clinical Signs and Symptoms Compared With the Criterion Standard for Examination Days 1 and 2 Using Spearman Rank Coefficients^a

Table 3

Physician No.	C1 vs N Phys Trial 1			C1 vs N Phys Trial 2		
	Day 1		Day 2	Day 1		Day 2
	<i>r_s</i>	<i>P</i> Value	<i>r_s</i>	<i>P</i> Value	<i>r_s</i>	<i>P</i> Value
1	0.28	.19	0.29	.17	0.56	.004
2 ^b	0.42	.04	0.34	.10		
3	0.56	.005	0.48	.02	0.71	<.001
4	0.41	.046	0.40	.05	0.81	<.001
5	0.49	.01	0.38	.06	0.62	.001
6	0.45	.03	0.58	.003	0.58	.003
7	0.35	.09	0.54	.006	0.57	.004
8	0.35	.09	0.36	.09	0.62	.001
9	0.56	.005	0.53	.008	0.66	<.001
10	0.47	.02	0.24	.26	0.63	<.001
11	0.41	.049	0.48	.02	0.68	<.001
12	0.57	.004	0.51	.01	0.46	.02
13 ^b					0.65	<.001

^aIndicates the individual physicians' Spearman rank correlations (*r_s*) of sum scores of clinical signs and symptoms with the composite score of nerve conduction (NC) abnormality (calculated as the composite standard normal deviate score of 5 attributes of NC abnormality [Σ 5 NC nds]). Boldface type indicates statistically significant correlations. One-sided exact binomial tests assessing rates of significant correlation (H₀ [null hypothesis]: *P* = .5; H_a [alternative hypothesis]: *P* > .5) were not significant for day 1 (*P* = .07) or day 2 (*P* = .61) of Trial 1. The tests were significant for day 1 (*P* < .001) and day 2 (*P* < .001) of Trial 2. Exact McNemar tests comparing rates of significant correlation between Trials 1 and 2 were not significant for day 1 (*P* = .25) or for day 2 (*P* = .06).

^bPhysician 2 was unavailable for Trial 2 and was replaced by physician 13.